ARL-TR-7869 ● Nov 2016

US Army Research Laboratory

Evaluation of Visualization Tools for Computer Network Defense Analysts: Display Design, Methods, and Results for a User Study

by Christopher J Garneau and Robert F Erbacher

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer’s or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

ARL-TR-7869 ● Nov 2016

US Army Research Laboratory

Evaluation of Visualization Tools for Computer Network Defense Analysts: Display Design, Methods, and Results for a User Study

by Christopher J Garneau
Human Research and Engineering Directorate, ARL

Robert F Erbacher
Computational and Information Sciences Directorate, ARL

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)* November 2016	**2. REPORT TYPE** Technical Report	**3. DATES COVERED (From - To)** January 2013–September 2015

4. TITLE AND SUBTITLE Evaluation of Visualization Tools for Computer Network Defense Analysts: Display Design, Methods, and Results for a User Study	**5a. CONTRACT NUMBER**
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Christopher J Garneau and Robert F Erbacher	**5d. PROJECT NUMBER**
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) US Army Research Laboratory ATTN: RDRL-HRM-B Aberdeen Proving Ground, MD 21005-5425	**8. PERFORMING ORGANIZATION REPORT NUMBER** ARL-TR-7869
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)**
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Computer network defense (CND) analysts serve an increasingly vital role in the defense of our nation’s computing infrastructure. An important component of their work is the monitoring of suspicious activity identified by an intrusion detection system (IDS). While analysts are trained to quickly recognize abnormal patterns in textual log files, humans are generally not well suited for such processing in any large quantity. Many authors have proposed the use of visualization techniques to aid the cyber security analysts’ search activities; however, such techniques are not widely used by analysts. This report describes an evaluation of 2 graphical displays (a “parallel coordinates” display and a “node-link” display) compared against a traditional tabular arrangement with the goal of better understanding analyst performance and obtaining subjective feedback on the graphical alternatives. Both expert analysts and novices (students) participated in the study. Results show that analysts generally preferred familiar tools but were able to use some graphical alternatives (node-link) to achieve similar performance in less time. Students were not found to be effective surrogates for experienced analysts for research/validation of techniques. This report describes the development and design of the displays and the experiment, and provides insight into analyst needs and evidence on effective methods for validating cyber defense visualization tools based on results obtained.

15. SUBJECT TERMS

visualization, analysis tools, computer network defense (CND), cyber security, expert analysis

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Christopher J Garneau
a. REPORT Unclassified	**b. ABSTRACT** Unclassified	**c. THIS PAGE** Unclassified	UU	98	**19b. TELEPHONE NUMBER (Include area code)** 410-278-5814

Contents

List of Figures

Approved for public release; distribution is unlimited.

List of Tables

Acknowledgments

The authors would like to thank Renée Etoty for her valuable contributions to the preparation, execution, and publication of this research.

INTENTIONALLY LEFT BLANK.

1. Introduction

Computer network defense (CND) analysts serve an increasingly vital role in the defense of our nation's computing infrastructure. An important component of their work is the monitoring of suspicious activity identified by an intrusion detection system (IDS). IDS configurations generally err on the side of caution and so a great deal of activity identified as suspicious is indeed benign (false positives). Therefore, analysts monitoring such systems must quickly sift through vast amounts of data to identify truly suspicious activity that warrants further investigation. These data are typically presented as textual log files. While analysts are trained to quickly recognize abnormal patterns in these files, the human brain is generally not well suited for such processing in any large quantity. Many authors have proposed the use of visualization techniques to preprocess and graphically arrange data to aid the cyber security analysts' search activities; however, such techniques are not widely used by analysts. This report describes an evaluation of 2 graphical displays compared against a traditional tabular arrangement with the goal of better understanding analyst performance and obtaining subjective feedback on the graphical alternatives. Both expert analysts and novices (students) participated in the study. Two facets of the study are described in this report: 1) the design and implementation of the displays for use in the study as well as the data collection method and 2) results of the study and characteristics of the participants.

1.1 Visualization and Cyber Security

Information visualization may support computer security work in that it provides security analysts with better tools to discover patterns, detect anomalies, identify correlations, and communicate findings. Information visualization can be used for exploration discovery, decision making, and communication of complex ideas, and it helps deal with processing the influx of data. Ideally, any visualization tools should make use of interactivity whenever possible, allowing the user to adjust the display to gain a more meaningful understanding of the data being presented. Mapping the data spatially in a meaningful matter is the most important and challenging part for constructing effective visualizations (Kosara et al. 2001).

1.1.1 Existing Visualization Tools

Various workflow visualization tools have been proposed to help CND analysts with various tasks (e.g., identify salient features in data sets, track their analysis, reuse effective workflows, test hypotheses). Etoty et al. (2014) conducted a survey of 59 visualization tools and environments available within the cyber security domain that have been developed to aid the analyst with interpreting or making decisions within an IDS framework. The study found tools in various states of development (from predevelopment to production) and with various goals (including “monitoring”, “analysis”, and “response”). The study concluded that many available tools could meet analyst needs, but some needs remain unmet. See Etoty et al. (2014) for details on the various types of visualizations and tools.

1.1.2 Validation of Visualization Tools for CND Analysts

In general, the domain of cyber security has few empirical studies that validate the effectiveness of available visual displays. There are several possible factors that hinder evaluation and successful deployment of visualization tools for cyber security tasks: 1) the availability of expert analysts and realistic data sets for user studies in academia is very limited, 2) CND analysts are often reluctant to use or explore advanced display techniques, instead preferring tools used within their existing workflow, and 3) the use of students as substitutes for domain experts—common in academia—may lead to misleading conclusions (this assertion is explored in the current work).

Some studies have focused on analyst needs for dealing with network data or cyber security tasks. D’Amico et al. (2005a, 2005b) used cognitive task analysis (CTA) to identify analysts’ reasoning processes when identifying and remedying network intrusions. They sampled 40 network security analysts and analyzed their threat mediation tactics (threat identification, threat quantification, threat correlation, attacker profiling, response formulation, and sense-making execution). Erbacher et al. (2010) conducted an extensive CTA with several experienced analysts and developed a set of realistic scenarios to be used for visualization validation. Requirements and characteristics of next-generation cyber situational awareness visualizations were also compiled from expert feedback.

Goodall (2009) conducted a comparative study and highlighted the benefits of user testing in general for cyber security visualizations. It was demonstrated that user performance (for students) for packet capture and analysis was better with the visualization tool The Network Visualizer (TNV) than with the traditional tool Ethereal (Wireshark), used by domain experts.

These studies notwithstanding, more research to better understand analyst needs and validate visual tools will benefit the cyber security research community and yield better tools for analysts.

1.2 Goals and Scope

As discussed in the previous section, numerous visualization techniques have been developed over the past decade, but few studies have evaluated and compared visualization tools in this domain to determine how real-world analysts might respond to and use them. The goal of the study described in this report is to address this gap with a user study based on a cyber security analysis scenario. Specifically, the following questions are considered:

- Do graphical displays and visualizations for CND analysts scanning IDS alert data increase performance compared with a traditional tabular arrangement?
- Is the use of students as substitutes for real-world end users for user studies a valid technique?
- What barriers currently limit the adoption of graphical displays by experienced analysts?
- What types of graphical displays would be most effective for enhancing analyst performance?

The US Army Research Laboratory (ARL) is particularly well suited to examine these issues because of its team of experienced in-house analysts that are part of its operational computer network defense service provider (CNDSP). This study investigates interpretation of display components, quantitatively compares tabular versus graphical displays, and captures analyst perception of the displays via various subjective measures. An additional component of the study compares real-world analyst feedback with that of students at Morgan State University because students are the primary test subjects for many academic studies that have the goal of developing visual displays for network monitoring.

In the main task in this study, participants act as analysts and their job is to identify as many of the network threats as possible in the provided IDS alert data. Objective response variables include true positive and false positive rates of identification of intrusion attempts and the time required for identification; subjective feedback includes responses to structured and free response questions.

2. Methods, Assumptions, and Procedures

This section describes the design of the study, including the design and development of the 3 displays (1 tabular and 2 graphical).

2.1 Study Design

As mentioned in the introduction, the study collected participant information using several pre- and post-task questionnaires as well as objective and subjective feedback during the main task of analyzing IDS alert data. The questionnaires appear in Appendix A. The study is a within-subjects design. Each participant completed the task using each of the 3 displays (displays are described in Section 2.4). The participant was assigned a number between 0 and 5 indicating the order of presentation of the 3 displays. Possibilities were as follows:

- 0: Tabular → Parallel Coordinates → Node-Link
- 1: Parallel Coordinates → Tabular → Node-Link
- 2: Node Link → Tabular → Parallel Coordinates
- 3: Tabular → Node Link → Parallel Coordinates
- 4: Parallel Coordinates → Node Link → Tabular
- 5: Node Link → Parallel Coordinates → Tabular

These numbers were assigned to participants to keep a similar number of participants in each category. This randomization was intended to avoid effects of practice and order bias.

2.2 Study Procedure

Data for the study were collected using a web-based questionnaire environment (LimeSurvey[*]) and custom web-based visualizations (discussed in Section 2.4). Prior to beginning the study, use and key features of the 3 displays were demonstrated by the experimenters on a projector or large monitor. The sessions with expert analysts took place at Adelphi Laboratory Center and Aberdeen Proving Ground and were conducted in groups of 1 or 2 at a time in a small room, with each analyst completing the survey using an individual personal computer with an external monitor. Sessions with students were conducted using a similar setup at Morgan State University (MSU). The study was approved by an appropriately constituted institutional review board at ARL and MSU (an early version of the protocol appears in Etoty et al. [2014]).

Participants were first briefed on the study and completed an informed consent form. As a participant began the study, they were instructed to complete several pretask questionnaires, including: 1) demographic information, 2) general

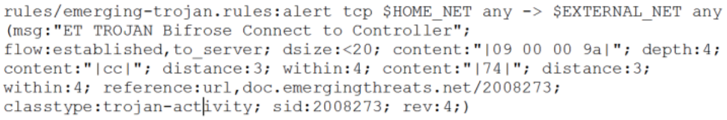
[*] https://www.limesurvey.org/en/

background information, and 3) a pretask analysis preference survey. Next, the participant was directed to open a custom web page that prompted the participant for their assigned display order (number 0 through 5). After this, they were taken to the first assigned display. For each display, participants were asked to group and select alerts constituting an intrusion attempt and provide comments as to their reasoning. Once they felt they identified all intrusion attempts, they submitted their results and progressed to the next display. After using all 3 displays, the participant completed several posttask questionnaires, including 1) a survey on usefulness and ease of use for each of the 3 displays, 2) an analysis survey on node representation, 3) an analysis survey on link representation, and 4) a posttask analysis preference survey. All questionnaires were implemented in LimeSurvey and appear in Appendix A.

2.3 Source Data for Displays

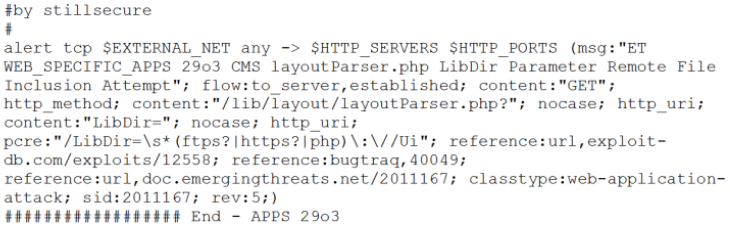
For the main task, participants were asked to analyze a set of synthesized alert data presented on the 3 displays. For security reasons, it was not possible to use data captured from ARL's network, so data were fabricated for the study. A description of how these data were created follows.

First, actual IDS data output was used, focusing only on an hour of alert messages and then parsing the data using Snort Rules.* Figures 1–4 show the example rule sets and our final parsed data set.



```
rules/emerging-trojan.rules:alert tcp $HOME_NET any -> $EXTERNAL_NET any
(msg:"ET TROJAN Bifrose Connect to Controller";
flow:established,to_server; dsize:<20; content:"|09 00 00 9a|"; depth:4;
content:"|cc|"; distance:3; within:4; content:"|74|"; distance:3;
within:4; reference:url,doc.emergingthreats.net/2008273;
classtype:trojan-activity; sid:2008273; rev:4;)
```

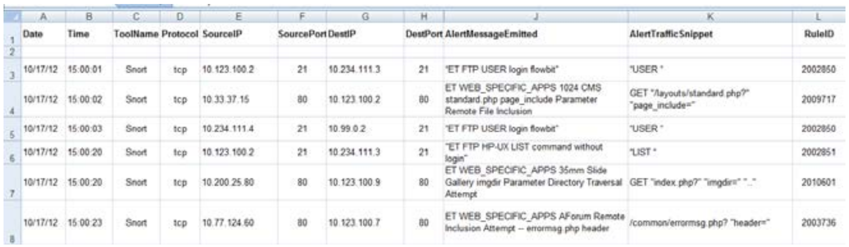
Fig. 1 Example rule set used for a typical threat named Bifrose from emerging trojan.rules and emerging-current _events.rules 0



```
#by stillsecure
#
alert tcp $EXTERNAL_NET any -> $HTTP_SERVERS $HTTP_PORTS (msg:"ET
WEB_SPECIFIC_APPS 29o3 CMS layoutParser.php LibDir Parameter Remote File
Inclusion Attempt"; flow:to_server,established; content:"GET";
http_method; content:"/lib/layout/layoutParser.php?"; nocase; http_uri;
content:"LibDir="; nocase; http_uri;
pcre:"/LibDir=\s*(ftps?|https?|php)\:\//Ui"; reference:url,exploit-
db.com/exploits/12558; reference:bugtraq,40049;
reference:url,doc.emergingthreats.net/2011167; classtype:web-application-
attack; sid:2011167; rev:5;)
################## End - APPS 29o3
```

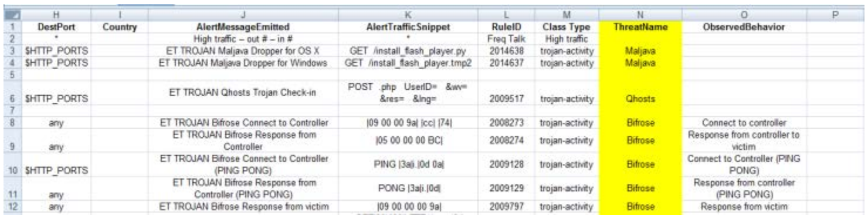
Fig. 2 Example rule set used for typical nonthreat traffic from emerging ftp.rules 0

[*] http://rules.emergingthreats.net/open/snort-2.9.0/rules/



	A	B	C	D	E	F	G	H	J	K	L
1	Date	Time	ToolName	Protocol	SourceIP	SourcePort	DestIP	DestPort	AlertMessageEmitted	AlertTrafficSnippet	RuleID
2											
3	10/17/12	15:00:01	Snort	tcp	10.123.100.2	21	10.234.111.3	21	"ET FTP USER login flowbit"	"USER "	2002850
4	10/17/12	15:00:02	Snort	tcp	10.33.37.15	80	10.123.100.2	80	ET WEB_SPECIFIC_APPS 1024 CMS standard.php page_include Parameter Remote File Inclusion	GET "/layouts/standard.php?" "page_include="	2009717
5	10/17/12	15:00:03	Snort	tcp	10.234.111.4	21	10.99.0.2	21	"ET FTP USER login flowbit"	"USER "	2002850
6	10/17/12	15:00:20	Snort	tcp	10.123.100.2	21	10.234.111.3	21	"ET FTP HP-UX LIST command without login"	"LIST "	2002851
7	10/17/12	15:00:20	Snort	tcp	10.200.25.80	80	10.123.100.9	80	ET WEB_SPECIFIC_APPS 35mm Slide Gallery imgdir Parameter Directory Traversal Attempt	GET "index.php?" "imgdir=" "..."	2010601
8	10/17/12	15:00:23	Snort	tcp	10.77.124.60	80	10.123.100.7	80	ET WEB_SPECIFIC_APPS AForum Remote Inclusion Attempt -- errormsg.php header	/common/errormsg.php? "header="	2003736

Fig. 3 Snippet of IDS alert data that was parsed using Open Snort 2.9.0 rules 0



	H	I	J	K	L	M	N	O	P
1	DestPort	Country	AlertMessageEmitted	AlertTrafficSnippet	RuleID	Class Type	ThreatName	ObservedBehavior	
2	*		High traffic – out # – in #	*	Freq Talk	High traffic			
3	$HTTP_PORTS		ET TROJAN Maljava Dropper for OS X	GET /install_flash_player.py	2014638	trojan-activity	Maljava		
4	$HTTP_PORTS		ET TROJAN Maljava Dropper for Windows	GET /install_flash_player.tmp2	2014637	trojan-activity	Maljava		
5									
6	$HTTP_PORTS		ET TROJAN Qhosts Trojan Check-in	POST .php UserID= &wv= &res= &lng=	2009517	trojan-activity	Qhosts		
7									
8	any		ET TROJAN Bifrose Connect to Controller	|09 00 00 9a| |cc| |74|	2008273	trojan-activity	Bifrose	Connect to controller	
9	any		ET TROJAN Bifrose Response from Controller	|05 00 00 00 BC|	2008274	trojan-activity	Bifrose	Response from controller to victim	
10	$HTTP_PORTS		ET TROJAN Bifrose Connect to Controller (PING PONG)	PING |3a|.|0d 0a|	2009128	trojan-activity	Bifrose	Connect to Controller (PING PONG)	
11	any		ET TROJAN Bifrose Response from Controller (PING PONG)	PONG |3a|i.|0d|	2009129	trojan-activity	Bifrose	Response from controller (PING PONG)	
12	any		ET TROJAN Bifrose Response from victim	|09 00 00 00 9a|	2009797	trojan-activity	Bifrose	Response from victim	

Fig. 4 Sample of completed table of parsed rules for the data set. Actual threats are highlighted in yellow.

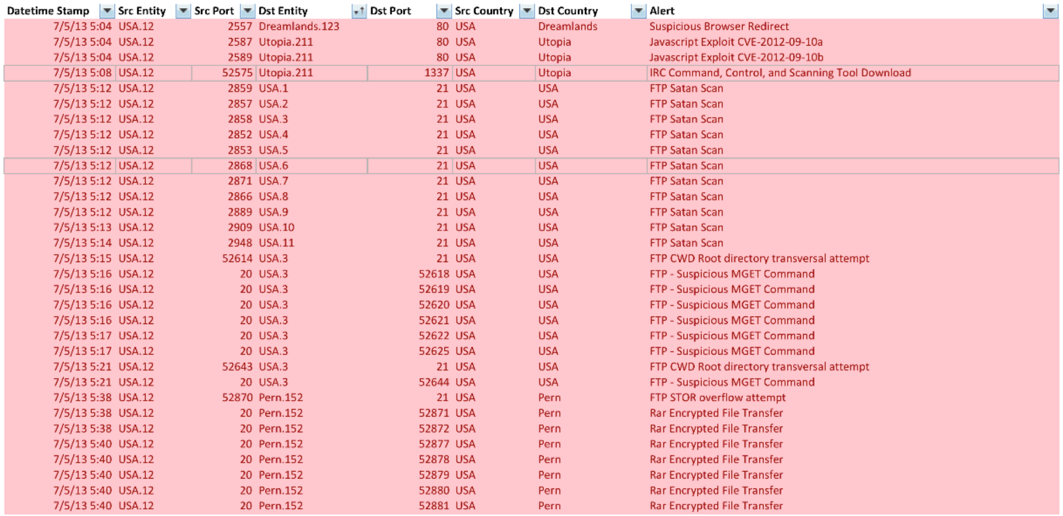
To compose a representative data set we defined parameters similar to the general look and feel of real-world "inbound and outbound" communications between the United States of America and other countries in a way that even a non-expert participant would understand. Normally host machines, computers, and other similar network-connected devices contain an assigned IP address that is the numeric address of their location when connected to the Internet. Thus for this study, we determined that a textual node identity, similar to a numeric IP address, would provide all the required linking information needed to define the topology of the scenario network. To keep the study simple and limit the effects of bias against any particular country, fantasylands are used as the node identities to represent foreign countries. These names included Land of Oz, Far Far Away, Neverland, Narnia, and others.*

The data set uses an attack scenario where many external nodes are attacking a smaller number of friendly peer nodes. The IP address for each node was selected from a synthetic grouping of addresses where the higher-order address octet values were common to a presumed "country" of origin. Node linkage was then accomplished by essentially testing out each attack and inserting proper IP addresses of the attacker and target nodes into mock "alert records" to constitute the scenario. Specifically, we used the IP address notation as the network identifier for each node in our fabricated network system data. The first 2 bytes were used to represent the country name (e.g., **United States**.Alabama.12). The

* http://listverse.com/2008/07/30/top-10-fantasy-worlds-in-literature/

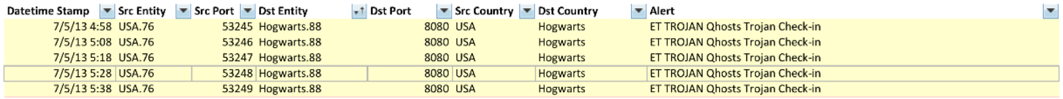
last 2 bytes represented the host identifier (e.g., United States.**Alabama.12**). In this example, Alabama would signify where an inbound connection within the United States was coming from, and the number 12 would be the label for the actual host machine (e.g., printer, fax, server, and computer).

The alert data contain 3 types of intrusion attempts of varying difficulty: 1) a 3-stage intrusion that consisted of a web infection, scanning, and data exfiltration; 2) periodic Trojan scanning; and 3) Sality Trojan infection. Figures 5–7 list the "ground truth" for each of these data sets. There were 139 total alerts with 42 alerts associated with 1 of the 3 intrusion attempts (true positives). Each record in the data contains 8 parameters that are presented in all the displays: 1) alert ID, 2) date/time stamp, 3) source entity (IP), 4) source port, 5) destination entity (IP), 6) destination port, 7) destination country, and 8) alert message. In addition to the true positive alerts, it was ensured that enough false positive alerts were present in the data (representing benign traffic) to provide sufficient "noise" in the alert data set.



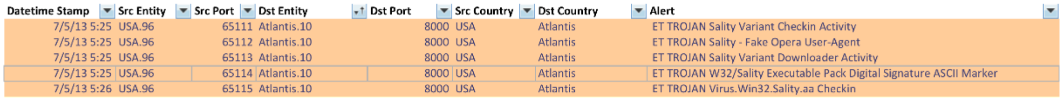
Datetime Stamp	Src Entity	Src Port	Dst Entity	Dst Port	Src Country	Dst Country	Alert
7/5/13 5:04	USA.12	2557	Dreamlands.123	80	USA	Dreamlands	Suspicious Browser Redirect
7/5/13 5:04	USA.12	2587	Utopia.211	80	USA	Utopia	Javascript Exploit CVE-2012-09-10a
7/5/13 5:04	USA.12	2589	Utopia.211	80	USA	Utopia	Javascript Exploit CVE-2012-09-10b
7/5/13 5:08	USA.12	52575	Utopia.211	1337	USA	Utopia	IRC Command, Control, and Scanning Tool Download
7/5/13 5:12	USA.12	2859	USA.1	21	USA	USA	FTP Satan Scan
7/5/13 5:12	USA.12	2857	USA.2	21	USA	USA	FTP Satan Scan
7/5/13 5:12	USA.12	2858	USA.3	21	USA	USA	FTP Satan Scan
7/5/13 5:12	USA.12	2852	USA.4	21	USA	USA	FTP Satan Scan
7/5/13 5:12	USA.12	2853	USA.5	21	USA	USA	FTP Satan Scan
7/5/13 5:12	USA.12	2868	USA.6	21	USA	USA	FTP Satan Scan
7/5/13 5:12	USA.12	2871	USA.7	21	USA	USA	FTP Satan Scan
7/5/13 5:12	USA.12	2866	USA.8	21	USA	USA	FTP Satan Scan
7/5/13 5:12	USA.12	2889	USA.9	21	USA	USA	FTP Satan Scan
7/5/13 5:13	USA.12	2909	USA.10	21	USA	USA	FTP Satan Scan
7/5/13 5:14	USA.12	2948	USA.11	21	USA	USA	FTP Satan Scan
7/5/13 5:15	USA.12	52614	USA.3	21	USA	USA	FTP CWD Root directory transversal attempt
7/5/13 5:16	USA.12	20	USA.3	52618	USA	USA	FTP - Suspicious MGET Command
7/5/13 5:16	USA.12	20	USA.3	52619	USA	USA	FTP - Suspicious MGET Command
7/5/13 5:16	USA.12	20	USA.3	52620	USA	USA	FTP - Suspicious MGET Command
7/5/13 5:16	USA.12	20	USA.3	52621	USA	USA	FTP - Suspicious MGET Command
7/5/13 5:17	USA.12	20	USA.3	52622	USA	USA	FTP - Suspicious MGET Command
7/5/13 5:17	USA.12	20	USA.3	52625	USA	USA	FTP - Suspicious MGET Command
7/5/13 5:21	USA.12	52643	USA.3	21	USA	USA	FTP CWD Root directory transversal attempt
7/5/13 5:21	USA.12	20	USA.3	52644	USA	USA	FTP - Suspicious MGET Command
7/5/13 5:38	USA.12	52870	Pern.152	21	USA	Pern	FTP STOR overflow attempt
7/5/13 5:38	USA.12	20	Pern.152	52871	USA	Pern	Rar Encrypted File Transfer
7/5/13 5:38	USA.12	20	Pern.152	52872	USA	Pern	Rar Encrypted File Transfer
7/5/13 5:40	USA.12	20	Pern.152	52877	USA	Pern	Rar Encrypted File Transfer
7/5/13 5:40	USA.12	20	Pern.152	52878	USA	Pern	Rar Encrypted File Transfer
7/5/13 5:40	USA.12	20	Pern.152	52879	USA	Pern	Rar Encrypted File Transfer
7/5/13 5:40	USA.12	20	Pern.152	52880	USA	Pern	Rar Encrypted File Transfer
7/5/13 5:40	USA.12	20	Pern.152	52881	USA	Pern	Rar Encrypted File Transfer

Fig. 5 The key (ground truth) for Intrusion Type 1 that was of an easy identification difficulty level. This was a multistage intrusion that consisted of a web infection onto a host, scanning to other machines, and finally an ftp exfiltration was done on one of the scanned host machines. This intrusion was successful.



Datetime Stamp	Src Entity	Src Port	Dst Entity	Dst Port	Src Country	Dst Country	Alert
7/5/13 4:58	USA.76	53245	Hogwarts.88	8080	USA	Hogwarts	ET TROJAN Qhosts Trojan Check-in
7/5/13 5:08	USA.76	53246	Hogwarts.88	8080	USA	Hogwarts	ET TROJAN Qhosts Trojan Check-in
7/5/13 5:18	USA.76	53247	Hogwarts.88	8080	USA	Hogwarts	ET TROJAN Qhosts Trojan Check-in
7/5/13 5:28	USA.76	53248	Hogwarts.88	8080	USA	Hogwarts	ET TROJAN Qhosts Trojan Check-in
7/5/13 5:38	USA.76	53249	Hogwarts.88	8080	USA	Hogwarts	ET TROJAN Qhosts Trojan Check-in

Fig. 6 The key (ground truth) for Intrusion Type 2 that was a medium identification difficulty level. There was a periodic Trojan checking in every 10 min for updates on a particular host machine called Hogwarts.88.



Datetime Stamp	Src Entity	Src Port	Dst Entity	Dst Port	Src Country	Dst Country	Alert
7/5/13 5:25	USA.96	65111	Atlantis.10	8000	USA	Atlantis	ET TROJAN Sality Variant Checkin Activity
7/5/13 5:25	USA.96	65112	Atlantis.10	8000	USA	Atlantis	ET TROJAN Sality - Fake Opera User-Agent
7/5/13 5:25	USA.96	65113	Atlantis.10	8000	USA	Atlantis	ET TROJAN Sality Variant Downloader Activity
7/5/13 5:25	USA.96	65114	Atlantis.10	8000	USA	Atlantis	ET TROJAN W32/Sality Executable Pack Digital Signature ASCII Marker
7/5/13 5:26	USA.96	65115	Atlantis.10	8000	USA	Atlantis	ET TROJAN Virus.Win32.Sality.aa Checkin

Fig. 7 The key (ground truth) for Intrusion Type 3 that was of hard identification difficulty level. This was a Sality Trojan infection.

2.4 Display Design

Three displays were chosen for evaluation in this study:

1) **Tabular display** (baseline): This display was intended to closely mirror existing analyst tools and was used as a baseline to which to compare the graphical displays. Basic functionality is similar to Microsoft Excel. See Section 2.4.2.

2) **Parallel coordinates display**: A "parallel coordinates" visualization was selected as 1 of the 2 graphical displays because it is one of the most published visualization techniques. Examples of parallel coordinates in the cyber security domain include implementations in the tools GeoViz (Giacobe and Xu 2011) and VIAssist (Goodall and Sowul 2009). See Section 2.4.3.

3) **Node-link display**: A "node-link" representation was selected as the second of the 2 graphical displays by an expert with experience as a CNDSP analyst. The display has been tailored to the task of intrusion detection based on related visualization research (Erbacher et al. 2002). See Section 2.4.4.

A description of the characteristics and functionality of each of these displays follows in the next sections. Each display was implemented as an interactive web interface. To expedite development, the tabular display and parallel coordinates displays were constructed by customizing available JavaScript libraries, and the node-link display makes use of raster graphics and JavaScript's "canvas" element. Details of the implementations are provided in the respective sections.

2.4.1 Common Elements

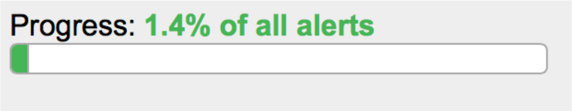
Participants were provided with 3 text prompts for each display: 1) "about", 2) "task", and 3) "instructions". The "about" and "task" prompts were the same for each display and are shown next; the "instructions" were unique for each display and are given with the display description in each respective subsection.

- **About:** "You are a cyber security analyst and your job is to use the provided IDS (Intrusion Detection System) tools to prevent or hamper future cyber attacks from countries against the United States of America (USA). You are viewing these items via a [display type] representation on a visual display."

- **Task:** "To make sure you are up for this mission we want you to complete the task of identifying intrusions or intrusion attempts from the enemy, which account for 25% of the total alert messages presented. If you reach 50% of the total alert messages presented, you will be directed to the next page. You have 20 minutes to complete this task."

For each of the 3 displays, several common interface elements are presented. They include the following:

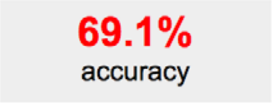
- Progress bar



Progress: 1.4% of all alerts

A progress bar indicates the percentage of the total available alerts that have been selected and submitted as an intrusion attempt. For instance, if a participant had identified 2 alerts as an intrusion attempt, the progress bar would show "1.4% of all alerts" (2/139=1.4%). If the participant reached a progress of 50%, they were forced to move on to the next display (this was intended to discourage mass selection of alerts).

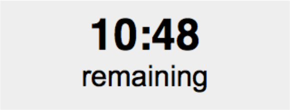
- Accuracy indicator



69.1%
accuracy

An accuracy indicator gives a sense of how a participant is performing. Accuracy is calculated as the number of true positives plus the number of true negatives divided by the total number of alerts (i.e., accuracy = (TP+TN)/ALL * 100%). If the participant submitted an intrusion attempt and the accuracy increased, the indicator turned green. If the accuracy decreased, the indicator turned red.

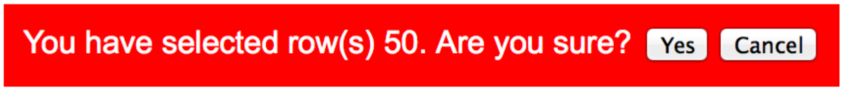
- Time remaining indicator



10:48
remaining

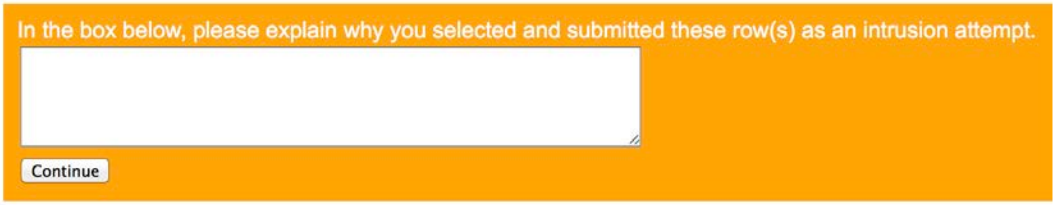
A timer gives the remaining time left in the session. To ensure timely completion of the study and simulate real-world time-pressure, a 20-min time limit per display was imposed. Once the timer reached 0, the participant was forced to move on to the next display.

- Submit Intrusion Attempt button and confirmation


You have selected row(s) 50. Are you sure?

A button allows the participant to submit 1 or more alerts as an intrusion attempt. A confirmation box (as shown) appears when the button is clicked.

- Comment box


In the box below, please explain why you selected and submitted these row(s) as an intrusion attempt.

Once the participant confirms their selection, a comment box appears to prompt the user to describe their rationale for selecting the alert(s).

- Pause Session/Session Complete buttons

A Pause Session button allowed the participant to take a break by darkening the screen, preventing any input, and stopping the timer. Once the Session Complete button was pressed, the user-identified intrusion attempts were saved and the participant was directed to the next assigned display.

As a participant moves from one display to the next, a custom PHP script saves a log file in the following format for the participant's performance on the given display:

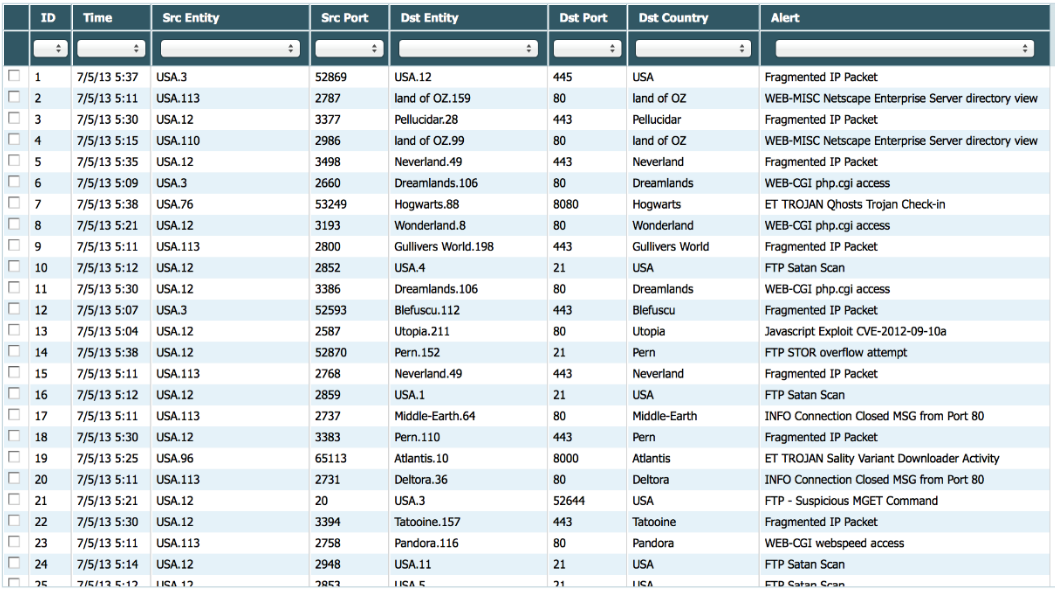
```
{"SUBMISSION_NUMBER_1": ["SELECTED_ID_1_1", "SELECTED_ID_1_2",
..., "SUBMISSION_1_COMMENTS", "SUBMISSION_1_TIME"],
"SUBMISSION_NUMBER_2": ["SELECTED_ID_2_1", "SELECTED_ID_2_2",
..., "SUBMISSION_2_COMMENTS", "SUBMISSION_2_TIME"], ...}

Time: start: OVERALL_START_TIME, end: OVERALL_END_TIME

Click Log: TIME_CLICK_1 @ (LOCATION_CLICK_1_X,
LOCATION_CLICK_1_Y); TIME_CLICK_2 @ (LOCATION_CLICK_2_X,
LOCATION_CLICK_2_Y); ...
```

2.4.2 Tabular Display

Figure 8 shows the tabular display presented to participants in the study. Analysts typically work with textual IDS log files, which may be either read as a text file or arranged in a spreadsheet. In general, displaying the data as a spreadsheet allows for more interactivity (filtering and sorting were the specific methods of interaction identified as being important for this study). The tabular display for this study was constructed with this functionality in mind by customizing the dhtmlxGrid Javascript library*. Appendix B, Section B-2, gives detailed pseudocode for this display. The dhtmlxGrid library has the ability to read and display a comma-separated value (CSV) file, which made transition of the data from Excel to the web interface straightforward.



	ID	Time	Src Entity	Src Port	Dst Entity	Dst Port	Dst Country	Alert
□	1	7/5/13 5:37	USA.3	52869	USA.12	445	USA	Fragmented IP Packet
□	2	7/5/13 5:11	USA.113	2787	land of OZ.159	80	land of OZ	WEB-MISC Netscape Enterprise Server directory view
□	3	7/5/13 5:30	USA.12	3377	Pellucidar.28	443	Pellucidar	Fragmented IP Packet
□	4	7/5/13 5:15	USA.110	2986	land of OZ.99	80	land of OZ	WEB-MISC Netscape Enterprise Server directory view
□	5	7/5/13 5:35	USA.12	3498	Neverland.49	443	Neverland	Fragmented IP Packet
□	6	7/5/13 5:09	USA.3	2660	Dreamlands.106	80	Dreamlands	WEB-CGI php.cgi access
□	7	7/5/13 5:38	USA.76	53249	Hogwarts.88	8080	Hogwarts	ET TROJAN Qhosts Trojan Check-in
□	8	7/5/13 5:21	USA.12	3193	Wonderland.8	80	Wonderland	WEB-CGI php.cgi access
□	9	7/5/13 5:11	USA.113	2800	Gullivers World.198	443	Gullivers World	Fragmented IP Packet
□	10	7/5/13 5:12	USA.12	2852	USA.4	21	USA	FTP Satan Scan
□	11	7/5/13 5:30	USA.12	3386	Dreamlands.106	80	Dreamlands	WEB-CGI php.cgi access
□	12	7/5/13 5:07	USA.3	52593	Blefuscu.112	443	Blefuscu	Fragmented IP Packet
□	13	7/5/13 5:04	USA.12	2587	Utopia.211	80	Utopia	Javascript Exploit CVE-2012-09-10a
□	14	7/5/13 5:38	USA.12	52870	Pern.152	21	Pern	FTP STOR overflow attempt
□	15	7/5/13 5:11	USA.113	2768	Neverland.49	443	Neverland	Fragmented IP Packet
□	16	7/5/13 5:12	USA.12	2859	USA.1	21	USA	FTP Satan Scan
□	17	7/5/13 5:11	USA.113	2737	Middle-Earth.64	80	Middle-Earth	INFO Connection Closed MSG from Port 80
□	18	7/5/13 5:30	USA.12	3383	Pern.110	443	Pern	Fragmented IP Packet
□	19	7/5/13 5:25	USA.96	65113	Atlantis.10	8000	Atlantis	ET TROJAN Sality Variant Downloader Activity
□	20	7/5/13 5:11	USA.113	2731	Deltora.36	80	Deltora	INFO Connection Closed MSG from Port 80
□	21	7/5/13 5:21	USA.12	20	USA.3	52644	USA	FTP - Suspicious MGET Command
□	22	7/5/13 5:30	USA.12	3394	Tatooine.157	443	Tatooine	Fragmented IP Packet
□	23	7/5/13 5:11	USA.113	2758	Pandora.116	80	Pandora	WEB-CGI webspeed access
□	24	7/5/13 5:14	USA.12	2948	USA.11	21	USA	FTP Satan Scan
□	25	7/5/13 5:12	USA.12	2853	USA.5	21	USA	FTP Satan Scan

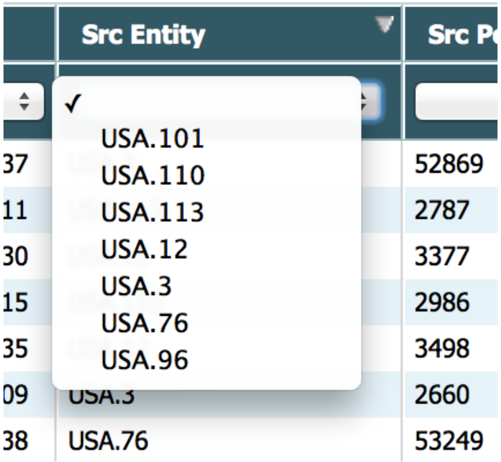
Fig. 8 Tabular display, showing alerts with ID 1-24

Instructions specific to this display (appearing under the “About” and “Task” prompts) were given as “Use the table to identify indicators of compromise by checking the box(es) in the associated row(s) of data. You may tap a column heading to sort data by the values in that column. You may also select a value from the drop-down box in each column to filter data. When you have selected the indicators associated with an intrusion attempt, click the Submit Intrusion Attempt button.

As a participant explored the data, they could click on a column heading to sort, or select an entry from the drop-down boxes to filter, as shown in Fig. 9. A

[*] http://dhtmlx.com/docs/products/dhtmlxGrid/

limited number of rows (about 25) were displayed at a given time due to screen size limitations, so the participant had to scroll to view the entire data set. To identify rows of interest for association as an intrusion attempt, the participant had to click on the checkbox at the start of the row. Once a row or rows had been selected as an intrusion attempt, they were grayed out and could not be selected again.



	Src Entity	Src P
37		52869
11		2787
30		3377
15		2986
35		3498
09	USA.3	2660
38	USA.76	53249

Fig. 9 Snippet of tabular display showing sorting and filtering capability

2.4.3 Parallel Coordinates Display

Figure 10 shows the parallel coordinates display presented to participants in the study. In general, a parallel coordinates display plots traces that pass through a single point on each vertical axis. Figure 11 shows a simple parallel coordinates display for 2 traces/records (orange and blue) on 3 parameters (x, y, and z). Units and scaling on each axis may differ and each axis may display different types of data (e.g., continuous, categorical). One strength of the display is that it allows its user to compare and contrast values of variables for large multivariate data sets to observe trends in the data.

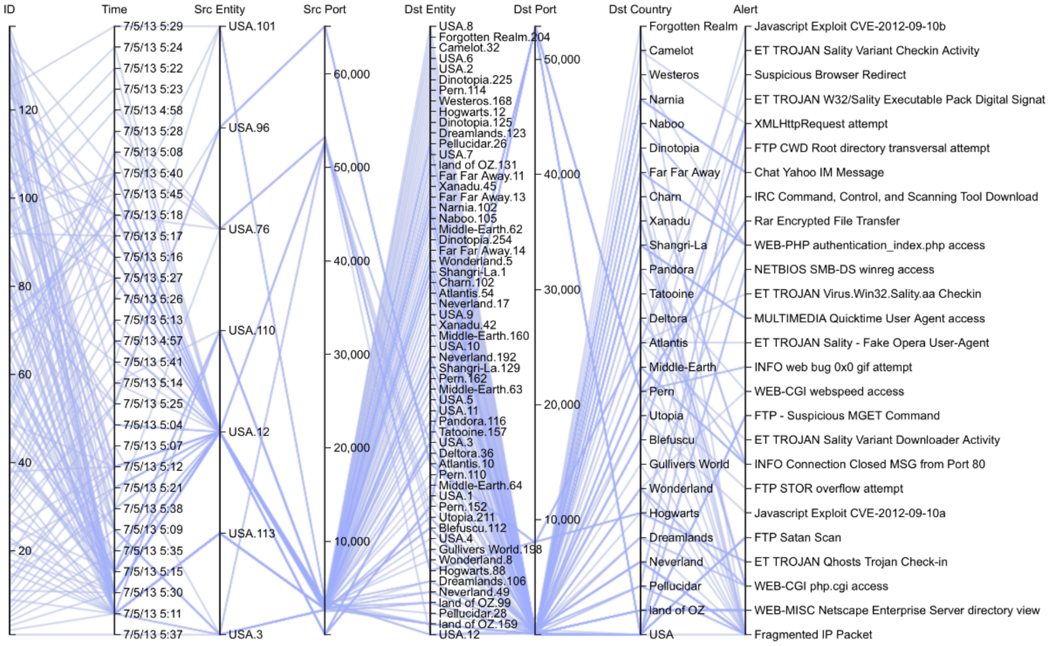


Fig. 10 Parallel coordinates display shown to participants (no alerts selected)

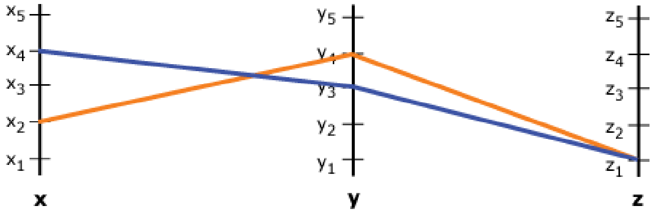


Fig. 11 Example of a generalized parallel coordinates display with 2 records and 3 parameters

For this study, each record in the alert data set constitutes a trace and passes through a value on each of the parallel axes (the 8 available parameters). The parallel coordinates display as implemented in the study is based on the Parallel Coordinates toolkit for D3.js.* Appendix B, Section B-3, gives detailed pseudocode for this display and demonstrates how the display was modified to meet the needs of the study (e.g., dimming selected traces). Like the dhtmlxGrid library, the D3.js parallel coordinates library has the ability to read and display a CSV file.

Instructions specific to this display (appearing under the "About" and "Task" prompts) were given as "Use the figure to identify indicators of compromise. You may filter and select data by highlighting a range of values on one or more axes.

* https://syntagmatic.github.io/parallel-coordinates/

When you have selected the indicators associated with an intrusion attempt, click the Submit Intrusion Attempt button.

To select alerts, the participant had to highlight/filter a range of values on any of the axes. To further refine the selection, ranges on additional axes could be selected. Figure 12 shows an example of a highlighted range used to select traces. To reset the filters, the participant could click a Reset Filters button at the bottom of the screen or click anywhere on the axes outside of the highlighted range. Once a set of traces had been selected and submitted as an intrusion attempt, the traces were grayed out and could not be selected again.

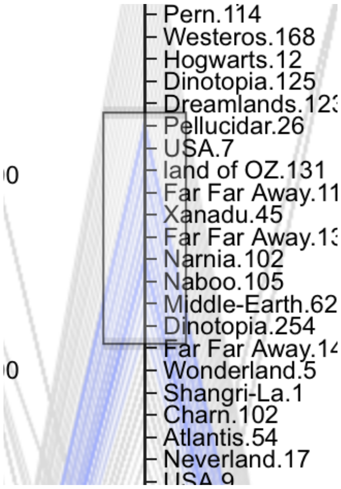


Fig. 12 Example of a highlighted range on the destination entity axis on the parallel coordinates display

2.4.4 Node-Link Display

Figure 13 shows the node-link display presented to participants in the study. This display was inspired by the glyph-based design previously used to visualize network traffic (Erbacher et al. 2002). In addition to displaying node labels on the nodes, the following features were used to encode alert data in the design:

- Line thickness of links, indicating number of alerts having the same pair of source node and destination node.
- Arrows of links, indicating direction of data communication.
- Spatial arrangement of nodes, indicating internal (bottom) and external (top) nodes.

- Configuration and color of source node rings: the total time was divided into 5-min increments shown as rings, and the rings were colored based on time of alert occurrence (color became lighter as time increased).

Instructions specific to this display (appearing under the "About" and "Task" prompts) were given as "Use the image to identify indicators of compromise by selecting markers corresponding with links of interest. You may click multiple markers. Hovering over the markers gives details about alerts associated with the link. When you have selected the indicators associated with an intrusion attempt, click the Submit Intrusion Attempt button.

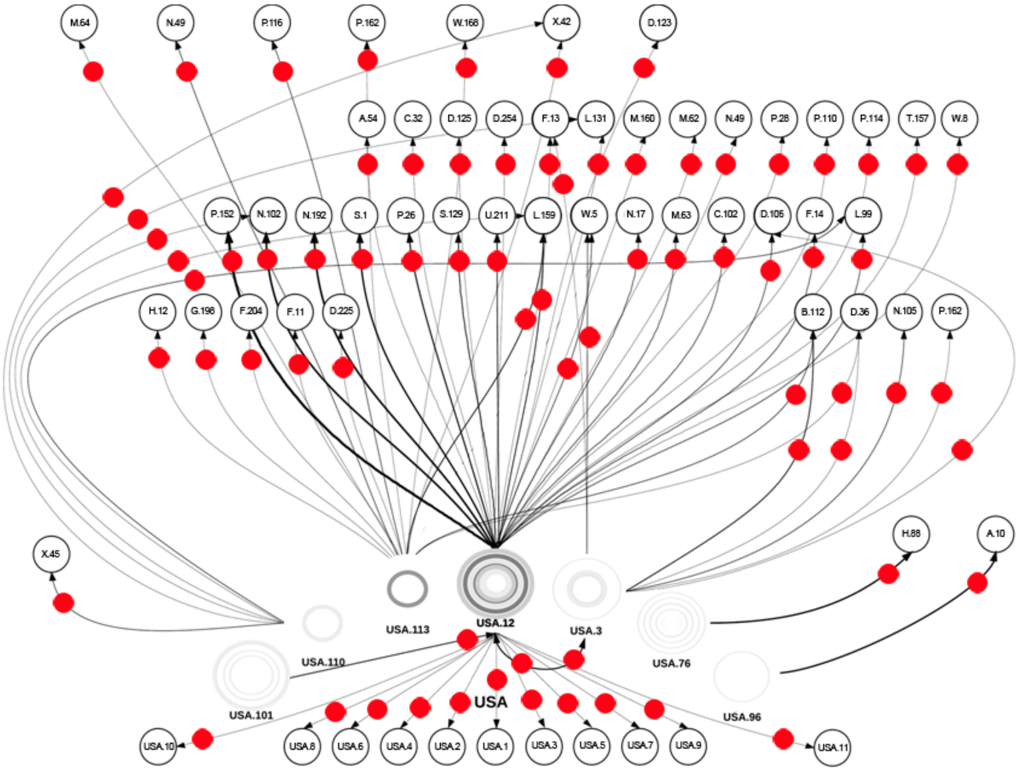
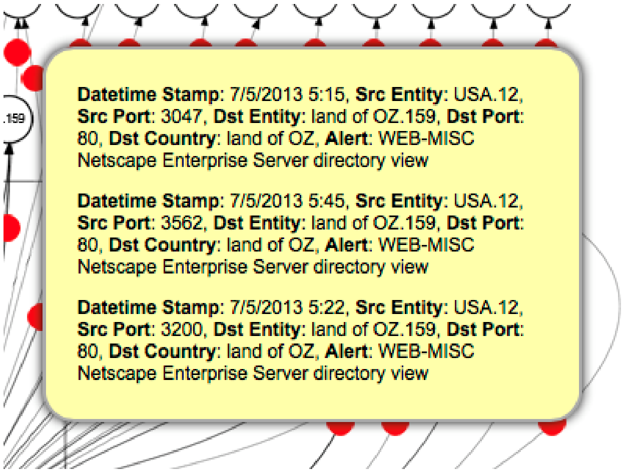


Fig. 13 Node-link display (no alerts selected)

The display was implemented as a layer of clickable red markers superimposed over a raster background image showing nodes and links. The background image was created by drawing 8 sections via Lucidchart* and then superimposing them to make 1 visual representation. The interactive clickable markers were achieved using JavaScript's canvas element.

*http://www.lucidchart.com/

As a participant moved the mouse over a marker, a popup appeared showing details of the alert associated with the source-destination node pair. Figure 14 shows an example of an alert popup. The color of the markers changed as the user clicked (selected) a marker; Fig. 15 shows an actively selected marker and Fig. 16 shows a marker that had been submitted. Once submitted, the participant could not select that marker again. Appendix B, Section B-4, gives detailed pseudocode for this display.



Datetime Stamp: 7/5/2013 5:15, **Src Entity**: USA.12, **Src Port**: 3047, **Dst Entity**: land of OZ.159, **Dst Port**: 80, **Dst Country**: land of OZ, **Alert**: WEB-MISC Netscape Enterprise Server directory view

Datetime Stamp: 7/5/2013 5:45, **Src Entity**: USA.12, **Src Port**: 3562, **Dst Entity**: land of OZ.159, **Dst Port**: 80, **Dst Country**: land of OZ, **Alert**: WEB-MISC Netscape Enterprise Server directory view

Datetime Stamp: 7/5/2013 5:22, **Src Entity**: USA.12, **Src Port**: 3200, **Dst Entity**: land of OZ.159, **Dst Port**: 80, **Dst Country**: land of OZ, **Alert**: WEB-MISC Netscape Enterprise Server directory view

Fig. 14 Popup box that appears as user moves the mouse over a marker. All alerts associated with that source-destination node pair are shown.

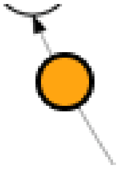


Fig. 15 Actively selected node-link marker

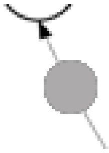


Fig. 16 Submitted node-link marker

3. Results

This section describes the results of the study, including the characteristics (demographics) of both cohorts of subjects (experts and students), objective performance while completing the main task of analyzing data using the 3 displays, and subjective feedback obtained via questionnaires.

3.1 Participant Characteristics

The participant population consisted of analysts from ARL and students from MSU. The analysts (experts) actively or previously had conducted CND analyses and were employed by ARL at the time of the study. They were assumed to be familiar with tabular tools and may or may not have been familiar with alternative graphical tools. Student participants (novices) were enrolled in courses in the Engineering and Computer Science programs at MSU at the time of the study. Students had either no experience or limited experience with CND analyses. There were 51 participants for the study from both groups combined; 24 subjects participated in the analyst cohort and 27 subjects participated in the student cohort. All participants were 18 years or older. All participants in both cohorts indicated that they had 20/20 vision (or corrected 20/20 vision), all passed a test for colorblindness, and none reported having any other disabilities. Note that none of the demographic questions forced a response, so some participants chose not to answer particular questions. Table 1 summarizes participant responses to the questions on demographics.

Table 1 Demographics for experts (analysts) and novices (students). Responses shown indicate choices available to participants ("other" was not an option unless otherwise indicated). Some participants did not answer one or more of these questions (responses were not required).

Demographic Characteristic	Experts (analysts)	Novices (students)
Gender		
Female	0	6
Male	22	17
Race		
White	13	2
Black or African American	5	22
Asian	1	2
Other	3	1
Age		
18–25 years	3	23
26–35 years	11	3
36–45 years	8	0
46–55 years	2	0
Highest level of education completed		
Some college but did not finish	5	NA
Two-year college degree/AA/AS	7	
Four-year college degree/BA/BS	7	
Some graduate work	3	
Completed Masters or professional degree	2	

3.2 Objective Performance

Objective performance is primarily measured in terms of 1) true positives (TPs), false positives (FPs), true negatives (TNs), and false negatives (FNs) that analysts identified in the alert data set; 2) measures derived from total TP, FP, TN, and FN identification, such as accuracy; and 3) time/duration to complete tasks. A response is considered to be a "true positive" if the alert identified by the participant was indeed an intrusion attempt when compared with ground truth; a "false positive" is any response identified as an intrusion attempt by the participant that is a benign alert when compared with ground truth; a "true negative" is any alert that was not identified by the participant and is a benign alert (correctly rejected); and a "false negative" is any response not identified as an intrusion attempt by the participant that is in fact an intrusion attempt when compared with ground truth (incorrectly rejected). Accuracy (A), precision (P), and recall (R) are defined in terms of the total TP, FP, TN, and FN identified as follows:

$$A = \frac{TP + TN}{TP + FP + FN + TN}$$

$$P = \frac{TP}{TP + FP}$$

$$R = \frac{TP}{TP + FN}$$

Objective performance results for each cohort are described in the next 2 subsections. The values given were calculated from the data log files described in Section 2.4.1 using a custom R script (given in Appendix C).

3.2.1 Objective Performance for Expert/Analyst Cohort

A summary of objective performance measures for the analyst cohort is given in Table 2. Note that not all analysts completed the task for all displays—2 analysts completed the task for the node-link and tabular but not parallel coordinates, 1 analyst completed task for the tabular but not node-link or parallel coordinates, and 1 analyst completed the task for the node-link and parallel coordinates but not the tabular.

Table 2 Objective performance parameters for each of the 3 displays for the analyst cohort. TP and FP give the average number of true positives and false positives identified, respectively. Twenty-four analysts participated in this cohort; n indicates the number of responses considered for the given display—this metric varies because not all analysts completed the task using all displays.

Parameter	Tabular	Parallel coordinates	Node-link
n	23	21	23
TP	24.8	20.3	25.9
FP	17.1	30.4	15.7
Completion time (min)	15.6	11.9	12.2
Accuracy	0.752	0.624	0.771
Precision	0.670	0.501	0.703
Recall	0.590	0.482	0.617

3.2.2 Objective Performance for Novice/Student Cohort

A summary of objective performance measures for the student cohort is given in Table 3. Note that not all students completed the task for all displays—1 student completed the task for the node-link and tabular but not parallel coordinates, 1 student completed the task for the parallel coordinates and tabular but not the node-link, and 1 student completed the task for the node-link and parallel coordinates but not the tabular.

Table 3 Objective performance parameters for each of the three displays for the student cohort. TP and FP give the average number of true positives and false positives identified, respectively. Twenty-seven students participated in this cohort; n indicates the number of responses considered for the given display—this metric is less than the total number of students because not all students completed the task using all displays.

Parameter	Tabular	Parallel coordinates	Node-link
n	26	26	26
TP	23.0	22.8	22.8
FP	30.5	35.2	29.6
Completion time (min)	12.7	9.3	7.9
Accuracy	0.652	0.616	0.656
Precision	0.489	0.458	0.49
Recall	0.561	0.557	0.556

3.2.3 Performance Comparison and Statistical Significance

The differences between means of various objective performance parameters are tabulated in Tables 4 and 5. Statistical significance is also indicated in the tables (2-tailed t tests assuming equal variance were conducted using R for each combination of parameters shown).

Table 4 Differences between means of various objective performance parameters are indicated for tabular vs. parallel coordinates (PC) and tabular vs. node-link. Positive values indicate higher values for the first of each pair (i.e., tabular display for tabular vs. PC and tabular vs. node-link, and node-link for node-link vs. PC). Significance is also indicated (* = Significant at the 0.05 probability level).

Parameter	Tabular vs. PC		Tabular vs. node-link		Node-link vs. PC	
	Expert	Novice	Expert	Novice	Expert	Novice
TP	+4.50	+0.19	-1.09	+0.27	+5.58	-0.077
FP	-13.34	-4.73	+1.35	+0.92	-14.7*	-5.65
Completion time (min)	+3.76*	+3.41*	+3.44*	+4.77*	+0.32	-1.36
Accuracy	+0.13*	+0.036	-0.019	-0.0042	+0.15*	+0.040
Precision	+0.17	+0.030	-0.033	-0.0015	+0.20*	+0.032
Recall	+0.11	+0.0046	-0.028	+0.0054	+0.14	-0.00077

Table 5 Differences between means of expert and novice performance for the objective performance parameters indicated. Positive values indicate higher values for the expert. Significance is also indicated (* = Significant at the 0.05 probability level).

Parameter	Tabular	PC	Node-link
TP	+1.74	-2.56	+3.10
FP	-13.41*	-4.80	-13.84*
Completion time (min)	+2.97*	+2.62	+4.30*
Accuracy	+0.10*	+0.0080	+0.12*
Precision	+0.18*	+0.043	+0.21*
Recall	+0.028	-0.074	+0.062

3.3 Subjective Feedback

Subjective feedback was collected via the described questionnaires. This section highlights responses to a few of the most salient subjective questions, which are organized by 1) structured questions and 2) free response questions.

3.3.1 Responses to Structured Questions

Tables 6 and 7 summarize the results for analysts and students, respectively, for the question for both the pre- and posttask survey: How much do you like these displays depending on their potential effectiveness of use? While the pretask questionnaires were administered before the participants used the displays for the analysis activity, participants had seen each of the 3 displays in the demonstration prior to beginning the study. The intent of asking the same question both before and after the main task was to assess how use of the displays may have changed participants' opinions about tabular and graphical representations.

Table 6 Pre- and posttask results for analysts for the question: How much do you like these displays depending on their potential effectiveness of use? n indicates number of responses to the question.

Display	Pre/Post	n	1	2	3	4	5	Average
Tabular	Pre	19	0	0	1	9	9	4.42
	Post	19	0	0	2	8	9	4.37
Parallel coordinates	Pre	18	3	2	8	2	3	3.0
	Post	19	10	3	1	4	1	2.11
Node-link	Pre	17	4	1	6	4	2	2.94
	Post	19	5	3	4	6	1	2.74

Notes: 1 = strongly dislike, 2 = somewhat dislike, 3 = neutral, 4 = somewhat like, 5 = strongly like.

Table 7 Pre- and posttask results for students for the question: How much do you like these displays depending on their potential effectiveness of use? n indicates number of responses to the question.

Display	Pre/Post	n	1	2	3	4	5	Average
Tabular	Pre	27	1	0	10	7	9	3.85
	Post	15	1	1	2	5	6	3.93
Parallel coordinates	Pre	27	1	0	8	9	9	3.93
	Post	15	0	1	3	8	3	3.87
Node-link	Pre	27	3	0	8	6	10	3.74
	Post	15	2	1	3	6	3	3.47

Notes: 1 = strongly dislike, 2 = somewhat dislike, 3 = neutral, 4 = somewhat like, 5 = strongly like

The posttask surveys were intended to assess the participants' experience with the displays. Table 8 summarizes analysts' perception of usefulness and appearance, and Table 9 summarizes how analysts compare and rate the types of displays. Tables 10 and 11 report results for the same questions for students.

Table 8 Analyst responses to subjective questions assessing usefulness and appearance. n indicates number of responses to the question.

Question	n	1	2	3	4	Average
"Overall, how would you rate the **usefulness** of the graphical layouts?"	24	8	9	7	0	1.96
"Overall, how would you rate the **usefulness** of the tabular layout?"	19	0	2	13	4	3.11
"Overall, how would you rate the **appearance** of the graphical layouts?"	19	8	8	1	2	1.84
"Overall, how would you rate the **appearance** of the tabular layout?"	23	0	3	14	6	3.13

Notes: 1 = poor, 2 = fair, 3 = good, 4 = excellent.

Table 9 Analyst responses to subjective questions indicated. n indicates number of responses to the question.

Question	n	1	2	3	4	5	Average
"I was able to complete my tasks better with the tabular display than the graphical displays."	24	0	1	4	3	16	4.42
"I prefer the graphical displays to the tabular display."	24	12	5	4	2	1	1.96
"I recommend that the use of the graphical displays be incorporated into analyst's cyber-security systems."	24	7	3	9	3	2	2.58
"I do not recommend that the graphical displays be incorporated into analyst's cyber-security systems."	24	3	5	9	2	5	3.04
"I recommend that the use of the tabular display be incorporated into analyst's cyber-security systems."	24	0	1	2	12	9	4.21
"I do not recommend that the tabular displays be incorporated into analyst's cyber-security systems."	24	11	7	6	0	0	1.79
"I easily understood the visualization of the graphical displays."	24	6	9	4	4	1	2.38
"I easily understood the visualization of the tabular display."	24	1	0	1	5	17	4.54
"The manipulation of the visualization's features of the graphical displays was easy."	24	6	7	5	4	2	2.54
"The manipulation of the visualization's features of the tabular display was easy."	24	0	0	2	5	17	4.63

Notes: 1 = strongly disagree, 2 = somewhat disagree, 3 = neutral, 4 = somewhat agree, 5 = strongly agree.

Table 10 Student responses to subjective questions assessing usefulness and appearance. n indicates number of responses to the question.

Question	n	1	2	3	4	Average
"Overall, how would you rate the **usefulness** of the graphical layouts?"	15	0	2	11	2	3.0
"Overall, how would you rate the **usefulness** of the tabular layout?"	14	0	0	13	1	3.07
"Overall, how would you rate the **appearance** of the graphical layouts?"	15	0	2	8	5	3.2
"Overall, how would you rate the **appearance** of the tabular layout?"	15	0	6	7	2	2.73

Notes: 1 = poor, 2 = fair, 3 = good, 4 = excellent.

Table 11 Student responses to subjective questions indicated. n indicates number of responses to the question.

Question	n	1	2	3	4	5	Average
"I was able to complete my tasks better with the tabular display than the graphical displays."	14	0	1	3	6	4	3.93
"I prefer the graphical displays to the tabular display."	14	1	1	3	7	2	3.57
"I recommend that the use of the graphical displays be incorporated into analyst's cyber-security systems."	14	0	2	1	6	5	4.0
"I do not recommend that the graphical displays be incorporated into analyst's cyber-security systems."	14	4	5	1	1	3	2.57
"I recommend that the use of the tabular display be incorporated into analyst's cyber-security systems."	14	0	0	4	4	6	4.14
"I do not recommend that the tabular displays be incorporated into analyst's cyber-security systems."	14	5	4	1	2	2	2.43
"I easily understood the visualization of the graphical displays."	14	0	0	3	6	5	4.14
"I easily understood the visualization of the tabular display."	15	0	1	3	4	7	4.13
"The manipulation of the visualization's features of the graphical displays was easy."	15	1	0	1	8	5	4.07
"The manipulation of the visualization's features of the tabular display was easy."	15	1	0	2	7	5	4.0

Notes: 1 = strongly disagree, 2 = somewhat disagree, 3 = neutral, 4 = somewhat agree, 5 = strongly agree.

3.3.2 Posttask Free Response

Feedback (both positive and negative) was obtained from the open-ended questions asked on the posttask questionnaires. Appendix D lists all responses to all open-ended questions (duplicate responses from a given participant and responses of less than 3 words are omitted and do not appear in the results). A sample of representative comments is highlighted in this section.

3.3.2.1 Analysts

“What components of the visual displays (table, parallel coordinates, and node link display) were most effective?”

- “Table is much better at identifying the actual alerts. Parallel is more involved with showing all the alerts that are on the node. Node Link is best at showing all the alerts that are attached to each ID”.
- “The table was the most effective. It is easy, you don't need all these fancy visual tools to find issues. Make everything simpler. The Node display and the Parallel display looked like a lot of noise. I wasn't interested in using them”.
- “I think the parallel coordinates display was the most helpful to me. I could quickly identify the most active areas. I think this would be a great tool to use in conjunction with the table display”.

“What aspects of the visualizations (table, parallel coordinates, and node link display) did you like best?”

- “I thought the node link display was interesting. It was nice to see a view of a network topology and the path the data travels”.
- “Visually separating the internal hosts from external hosts to quickly see flow of data between internal only and internal to external”.
- “Table is much better for seeing and lumping the alerts together. Parallel showed the flow a lot better. Node link showed the flow and other ID numbers to the alerts faster”.
- “I liked being able to quickly identify the most active times of the day and the most common request domains of the parallel coordinates displays”.

“What aspects of the visualizations (table, parallel coordinates, and node link display) did you not like?”

- “It takes a while to glance at everything when you can glance at percentages and numbers. Those are quicker to grasp sometimes.”
- “The line display and node display were convoluted at best; they detracted from the information presented. In terms of investigative procedure, I believe they would only serve to stifle”.
- “The parallel coordinates was a nightmare to visualize and the node link display was far too time consuming to hover through”.

- “Table it is hard to see the relation of the alerts harder, and if there are more than one alert type on the node it can be harder to see that. Parallel, has too much going on that it becomes really lost and confusing, this is definitely not a set up I would want on a larger network. I spend most of my time going through the network trying to make sure that I haven't skipped a node. Node link, has a good basis, but to search quickly it can be difficult if there are more ports added in. This version could use the nodes in it not just the IDs”.

- “Connections were very hard to follow, information was displayed in a non-intuitive manner, correlations were very difficult to find without excessive work”.

- “The parallel coordinates interface was not useful or intuitive and I could not sort the coordinates. Moreover, once I was finished with an alert it should have been removed from my view. Also I should be able to remove noise from my view with a filter. The node link display was slightly more useful but the node sizes were not intuitive nor were the most important piece of analysis data displayed up front: the alerts!”

- “Parallel coordinates is good for fine analysis but not for raw/bulk analysis”.

- “The graphical representations were completely unusable to me. The table was fine but there needs to be drill down options to see more data. I look at an alert then I check some traffic if I see something suspicious I dump the traffic and do a thorough investigation”.

“What features of the visual displays were most useful for completing your tasks in this study?”

- “Table data was fine but I need more than play data to do proper analysis. It isn't just the alert or a darker line of traffic that determines infection or compromise”.

“What did you learn from this study?”

- “Its good that I got a chance to see what type of tools can be deployed in future and felt very good to leave feedback about these tools”.

- “That there are some great relationship tools for network intrusion, and some not so great ones”.

- “Being able to see correlations is very important (time, src ip, dst ip, etc). All of this information is very difficult to fit into a graphical interface.

Without being able to sort and filter, this makes the analysts job far more difficult. The graphical displays seem decent for a "birds eye view", but a nightmare for actual everyday analysis".

- "Aggregated data obfuscates signal!"
- "A simple table can be the better option sometimes".
- "Graphics helps make analysis easier".
- "With all due respect regarding this project, it seems like there is a long way to go before a very useful graph or node visualization would be effective or efficient to use".
- "I do not like graphical displays while doing analysis. I find them highly unnecessary, unless of course I was presented with ones that are more intuitive".

3.3.2.2 Students

"What components of the visual displays (table, parallel coordinates, and node link display) were most effective?"

- "Table- the variety of filter options. Node-link- the simple graphics of it".
- "The unique way the data was portrayed gave the data an interesting yet effective way to show data".

"What aspects of the visualizations (table, parallel coordinates, and node link display) did you like best?"

- "The way the different components of the parallel coordinates connected and analyzed".

"What aspects of the visualizations (table, parallel coordinates, and node link display) did you not like?"

- "The parallel coordinates seemed to be a bit congested with the words".
- "The unsureness of what certain things meant i.e. thickness of links, etc."
- "The table was a bit bland in comparison to the other two displays".

"What aspects of the visualizations (table, parallel coordinates, and node link display) helped you to identify intrusions?"

- “Being able to see the amount of traffic and basic information in the node link display helped. Obviously the search options for the table were straight forward and helpful”.
- “Noticing clusters of alerts or intrusions”.
- “Looking at uncommon activity if it was particular names and times, or times and amount of activity from other networks, etc.”

“What features of the visual displays were most useful for completing your tasks in this study?”

- “The ability to sort and sift”.
- “The good organization of the data”.

“What did you learn from this study?”

- “Advancements in computer vision can greatly impact one's ability to perform a task. Not knowing much about what to look for in this exercise was not a hinderance”.
- “Risks can be categorized based on features that may exist. Network threats are real and something that is really trying to be prevented. Humans must being of some great importance over machines or of equivalent importance when it comes to analyzing threats, otherwise this study wouldn't be conducted”.
- “I learned a lot about how the cyber security field operates to try and locate potential computer hazards/viruses”.
- “How to analyze data and pick out abnormalities through a huge amount of data”.

4. Discussion

The performance and responses provided by analysts and students in this study provide a valuable opportunity to examine 1) the effectiveness of traditional tabular tools and graphical tools employing data visualization techniques for IDS alert analysis and 2) how display validation studies may be effectively carried out. This section is divided into 4 parts: 1) discussion of how analysts performed using the various types of displays, 2) discussion of how students performed compared with analysts and their suitability as a surrogate test population, 3) lessons learned from the study, and 4) future work.

4.1 Discussion of Analyst Performance

Analysts generally preferred and performed well on the tabular (baseline) display. This was to be expected, as analysts are most familiar with using this tool for alert analysis. Referencing Table 4 to compare the tabular display performance with the parallel coordinates display performance, analysts were faster using the parallel coordinate display but more accurate using the tabular display. Compared with the node-link display, analysts were faster using the node-link display than the tabular display and achieved about the same accuracy on both. Compared with the parallel coordinate display, analysts took about the same time on each but were more accurate using the node-link display. These differences are statistically significant except for the difference in accuracy between the tabular and node-link displays (for which analysts were similarly accurate) and the difference in time between the node-link and parallel coordinate displays. In Table 2, it is notable that analysts had a high rate of false positive identification for the parallel coordinates display. This would translate into more time wasted in a real-world analysis scenario investigating benign alerts as compared with the other 2 displays.

Subjective feedback is overwhelmingly positive for the tabular display and mixed for the other displays. While some analysts could see value in the parallel coordinates display, most felt strongly against it, exemplified by comments such as “parallel coordinates is a nightmare to visualize” and “the parallel coordinates interface was not useful or intuitive”. In general, analysts did not prefer the use of the graphical alternatives to the tabular display (Table 9 shows that analysts rated the response “I prefer the graphical displays to the tabular display” between “strongly disagree” and “somewhat disagree”.) One analyst commented, “The graphical representations were completely unusable to me.” However, other analysts found them useful. Of the node-link display, one analyst commented, “It was nice to see a view of a network topology and the path the data travels.” Another said, “Node link showed the flow and other ID numbers to the alerts faster.”

Section 4.3 further discusses the implications of these sentiments. While most analysts preferred the tabular display to graphical displays, the node-link display offered very similar performance (accuracy) to the tabular display while requiring less time to complete the task. Moreover, between the graphical displays, the analysts’ accuracy was significantly greater using the node-link display than the parallel coordinate display.

4.2 Discussion of Students as Surrogates

Objective performance and subjective feedback for the student cohort yield insight as to their effectiveness as surrogates in evaluations of display performance for CND analysts. Tables 3 and 4 show that students performed similarly using all 3 displays in terms of accuracy and rate of TP and FP identification. Tables 10 and 11 also showed that students felt similarly about the usefulness and appearance of both types of displays and felt that both were similarly easy to understand and manipulate. These results would indicate that a lack of contextual knowledge prevented students from effectively using any of the displays. This is confirmed by the comments provided to the free response questions—in many cases, students reported that they learned how to analyze data as a result of this study. When compared with expert analysts (Table 5), student performance on the tabular and node-link displays was significantly different (worse) in terms of false positive identification and accuracy (both experts and students performed poorly on the parallel coordinates display). Combined with the large differences in demographics between students and analysts (see Section 3.1), these results suggest that students are not effective surrogates for experienced CND analysts. However, the use of students as surrogates for analysts-in-training may be appropriate and future studies might investigate this scenario.

4.3 Lessons Learned from the Study

Reflecting on the development and execution of the study yields several "lessons learned" that may be applied to similar studies in the future. First, including as much interactivity as possible in the displays to be evaluated will greatly benefit the participant experience and thus enhance credibility of the results. Using hypertext markup language (HTML) and JavaScript as the development environment makes interactive display creation relatively easy and quick. Many freely available software libraries online greatly aided this task for this study (e.g., jQuery, D3.js). One analyst wrote, "Being able to filter out ports and hosts helped on the table." However, even more interactivity would have benefited the displays used in the study. One analyst commented, "The parallel coordinates interface was not useful or intuitive and I could not sort the coordinates. Moreover, once I was finished with an alert it should have been removed from my view." Another said, "Connections were very hard to follow, information was displayed in a non-intuitive manner, correlations were very difficult to find without excessive work." These problems could likely have been mitigated with even more interactivity embedded in the display design.

Additionally, other limitations in the study execution should be noted. First, in attempt to enhance participants' incentive to perform well (and lend a game-like quality to the test environment) an accuracy indicator was added to each of the displays (see Section 2.4.1). However, there are several drawbacks to its inclusion: it would not exist in a real-world scenario; it may influence participants' opinions about each tool; it may alter performance in unexpected ways (e.g., when the indicator turns red the analyst may feel prompted to slow down); and it is inaccurate in that the number of TNs—one of the components of accuracy—is not calculable until the participant completes the task for the display. Similarly, the timer adds a certain sense of realism and time pressure to the task, but the time limit value (20 min) was chosen arbitrarily and its benefit and effect on results is unclear.

At a high level, it is important to thoroughly explain the intent of the study to participants as well as the task to be performed. While specific instructions were provided to the analysts, some wanted more data and could not understand how to use the displays for the task provided. One said, "Table data was fine but I need more than play data to do proper analysis. It isn't just the alert or a darker line of traffic that determines infection or compromise." This view likely originated from a misunderstanding about the intent of the study. The displays were not intended to replicate a complete analysis session, but rather provide a tool for rapidly identifying indicators of compromise for further investigation. To design a complete environment (e.g., data of real-world appearance and scale, calls to threat cells or target sites that are fully scripted, and so on) would require years of research and design and may not provide results that justify such an undertaking. Leveraging expectations by noting to analysts that they were participating in a scaled-down study would have been beneficial.

Finally, if professional analysts are to be used in a validation study—which is ideal—it is important to emphasize possible future improvements to their workflow. While experienced analysts approach the study with a wealth of knowledge and insight, they have a certain way they approach their work and may be critical of alternatives. Some were quite receptive to new tools. For instance, one said, "It's good that I got a chance to see what type of tools can be deployed in future and felt very good to leave feedback about these tools." On the other hand, another had a more cynical viewpoint and responded, "The table was the most effective. It is easy, you don't need all these fancy visual tools to find issues. Make everything simpler. The Node display and the Parallel display looked like a lot of noise. I wasn't interested in using them." It is impossible—and undesirable—to eliminate contrary perspectives, but approaching the study with a "help me to help you" attitude may enhance results.

4.4 Future Work

There are many opportunities to build upon the results discussed in this report. First, the effect of specific elements of the visualizations (e.g., visual encoding, colors, size, spacing, and links) may be explored and optimized to increase effectiveness of both tabular and graphical displays. Detailed, practical recommendations for designing effective displays should be provided by examining results from this and/or similar studies. Modifying the experiment design by asking participants to self-rate confidence in their answers for comparison with actual accuracy scores could be insightful and might yield insight about the user experience of the interface. Future studies might better contextualize the visualizations within other tasks that analysts perform (i.e., alerts are only a part of understanding network traffic for cyber defense, and this was noted as a frustration by several analysts). Future studies might also investigate integrating elements of traditional/tabular and graphical displays (in some ways, the node-link display was a first attempt at fusing aspects of both).

5. Conclusion

This study provides an empirical evaluation of 3 display alternatives for cyber network defense analysts: a traditional tabular display, a graphical parallel coordinates display, and a graphical node-link display. Both experienced analysts and university students participated in the study. Results show that analysts generally preferred familiar tools but were able to use some graphical alternatives (node-link) to achieve similar accuracy in less time. Students were not found to be effective surrogates for experienced analysts for validation, calling into question a common practice. While the study has several limitations—including use of simulated data instead of real-world data as well as use of functionally limited custom displays instead of typical full-scale CND visualization tools—it provides insight into analyst needs and evidence on effective methods for validating cyber defense visualization tools.

6. References

D'Amico A, Whitley K, Tesone D, O'Brien B, Roth E. Achieving cyber defense situational awareness: a cognitive task analysis of information assurance analysts. Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting. 2005 Sep. Thousand Oaks (CA): SAGE Publications; 2005a;49(3):229–233.

D'Amico A, Kocka M. Information assurance visualizations for specific stages of situational awareness and intended uses: lessons learned. In: Ma K-L, North S, Yurcik B, editors. VizSEC 05. Proceedings of the IEEE Workshop on Visualization for Computer Security 2005; 2005 Oct 26; Minneapolis, MN. Piscataway (NJ): IEEE; c2005b. p. 107–112.

Erbacher R, Walker K, Frincke D. Intrusion and misuse detection in large-scale systems. IEEE Computer Graphics and Applications. 2002;22(1):38–47.

Erbacher R, Frincke D, Wong PC, Moody S, Fink G. A multi-phase network situational awareness cognitive task analysis. Information Visualization Journal. 2010;9(3):204–219.

Etoty RE, Erbacher R, Garneau C. Evaluation of the presentation of network data via visualization tools for network analysts. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2014 Mar. Report No.: ARL-TR-6865.

Giacobe N, Xu S. Geovisual analytics for cyber security: adopting the GeoViz toolkit. In: Miksch S, Ward M, editors. VAST 2011. Proceedings of the IEEE Conference on Visual Analytics Science and Technology 2011; 2011 Oct 23–28; Providence, RI. Piscataway (NJ): IEEE; c2011. p. 315–316.

Goodall JR, Sowul M. VIAssist: visual analytics for cyber defense. HST′09. Proceedings of the 2009 IEEE International Conference: Technologies for Homeland Security; 2009 May 11–12; Boston, MA. Piscataway (NJ): IEEE c2009. p. 143–150.

Goodall JR. Visualization is better! A comparative evaluation. In: Frincke DA, Gates CE, Goodall JR, Erbacher RF, editors. VizSec 2009. Proceedings of the 6th International Workshop on Visualization for Cyber Security; 2009 Oct 11; Atlantic City, NJ. Piscataway (NJ): IEEE; c2009. p. 57–68.

Kosara R, Miksch S, Hauser H. Semantic depth of field. In: Andrews K, Roth S, Wong PC, editors. INFOVIS′01. Proceedings of IEEE Symposium on Information Visualization 2001. INFOVIS; 2001 Oct 22–23; San Diego, CA. Piscataway (NJ): IEEE; c2001. p. 97–104.

Appendix A. Questionnaires Presented to Participants

This appendix appears in its original form, without editorial change.

This appendix presents the questionnaires displayed to participants. Responses available to participants are shown in brackets after each question.

Welcome to this study entitled “Evaluation of the Presentation of Network Data via Visualization Tools for Network Analysts”. You should receive an informed consent form and training by the study investigators. Please proceed to the next page when instructed to do so.

Demographic Information

1 What is the random ID assigned to you?
[Input box]

2 What is your gender?
Please choose **only one** of the following:
[Male, Female]

3 What is your race?
Please choose **only one** of the following:
[American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White, Other]

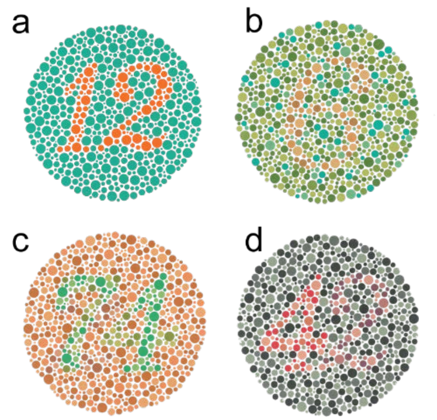
4 What is your age?
Please choose **only one** of the following:
[18-25, 26-35, 36-45, 46-55, 56-65, 66-75, 76 or older]

5 What is the highest level of education you have completed?
Please choose **only one** of the following:
[Elementary school only; Some high school, but did not finish; Completed high school; Some college, but did not finish; Two-year college degree / A.A / A.S.; Four-year college degree / B.A. / B.S.; Some graduate work; Completed Masters or professional degree; Advanced Graduate work / Ph.D.]

6 What is your work title?
[Input box]

7 What is your current department?
[Input box]

8 In the boxes below, enter the number you see for each letter (if any). If you do not see a number, enter "none".
[a, b, c, d; See image below]



9 Do you usually wear contacts or eye glasses for reading?
Please choose **only one** of the following:
[Yes, No]

10 Are you wearing your contacts or glasses right now?
Only answer this question if the following conditions are met:
Answer was 'Yes' at question '9 [DI9]' (Do you usually wear contacts or eye glasses for reading?)
Please choose **only one** of the following:
[Yes, No]

11 Do you have any other disabilities
Please choose **only one** of the following:
[No, Yes (please explain in the comment box), Input box]

General Background Information

12 Do you use computers (PC's, MAC, iPad, iPhone, Android phone, tablets, etc.)?
Please choose **only one** of the following:
[Yes, No]

13 How often on a daily basis, do you use computers?
Please choose **only one** of the following:
[1-5 hrs, 5-10 hrs, 10-15 hrs, 15-20 hrs, More than 20 hrs]

14 How comfortable do you feel using a computer?
Please choose **only one** of the following:
[Very comfortable, Somewhat comfortable, Somewhat uncomfortable, Very uncomfortable]

15 Have you ever written a software program or a mini computer code?
Please choose **only one** of the following:
[Yes, No]

16 Have you ever configured a Linux computer?
Please choose **only one** of the following:

[Yes, No]

17 What is a shell?

[Input box]

18 Have you ever worked as a network analyst or have any network analysis experience?
Please choose **only one** of the following:

[Yes (specify the experience, location, and time in the comment box), No, Input box]

19 How many years have you been a cyber analyst?
Only answer this question if the following conditions are met:
Answer was 'Yes (specify the experience, location, and time in the comment box)' at question '18 [7]' (Have you ever worked as a network analyst or have any network analysis experience?)
Please choose **only one** of the following:

[How many years have you been a cyber analyst?, Less than 1 year, 1 to 3 years, 3 to 5 years, 5 to 10 years, More than 10 years, Never]

20 Which of the following activities do you most frequently perform in your work?
Only answer this question if the following conditions are met:
Answer was 'Yes (specify the experience, location, and time in the comment box)' at question '18 [7]' (Have you ever worked as a network analyst or have any network analysis experience?)
Please choose **all** that apply:

[Filter raw sensor data (e.g. IDS alerts), Point out the suspicious activities from filtered data, Collect evidence from multiple sources (e.g. IDS, package dumps, etc.), Group individual activities and make hypotheses about an intrusion attempt, Assess attacker identity and mission impact, Tuning sensors to look for predicted attack, Incident handling, Produce documents to report current situation awareness, Perform virus/incident handling, Train others of situation awareness]

Pre-Task Questionnaire: Analysis Survey

21 On a scale from 1 to 5 where 1 is "Very Sad", 3 is "Neutral", and 5 is "Very Happy", which of the following best describes your current emotional state?
Please choose **only one** of the following:

[1: Very Sad, 2: Sad, 3: Neutral, 4: Happy, 5: Very Happy]

22 Which is better at the following tasks, Machine or Human?
Please choose the appropriate response for each item:
Analyzing data [Machine, Human]
Detecting anomalies (where an anomaly is abnormal behavior on a security network) [Machine, Human]

23 In general, which type of display do you find most useful in analyzing data?
Please choose **all** that apply:

[Tables, Textual, Line & Bar Graphs, Simple Graphs, Symbolic Shapes, Other (please specify)]

24 Please specify the other type of display you find useful in analyzing data.
Only answer this question if the following conditions are met:
Answer was at question '23 [3]' (In general, which type of display do you find most useful in analyzing data?)

[Input box]

25 If you are not a cyber security analyst, how interested are you in what they do?
Please choose **only one** of the following:

[Highly interested, Somewhat interested, Unsure, Not at all interested]

26 What motivated you to participate in this study?
Please write your answer here:

[Input box]

27 What do you expect to learn from this study?
Please write your answer here:

[Input box]

28 For the next two questions, recall the tabular, parallel coordinates, and node-link layouts from the earlier demo presented to you and respond on a scale from 1 to 5.

29 How effective do you believe these three different visual layouts will be in cyber analysis tasks?
Please choose the appropriate response for each item [1-5]:
Tabular
Parallel coordinates
Node-link

30 How much do you like these displays depending on their potential effectiveness of use?
Please choose the appropriate response for each item [1-Strongly dislike, 2-Somewhat dislike, 3-Neutral, 4-Somewhat like, 5-Strongly like]:

Tabular
Parallel coordinates
Node-link

Visual Displays Break

31 STOP. Please ask the test administrator to help with the next task.

Do not click the 'Next' button below until you have completed the task.

Post-Task Questionnaire, Usefulness and Ease of Use

32 The following questions assess "usefulness" and "ease of use" for each of the three displays.

33 For the TABULAR DISPLAY only, answer the following questions about usefulness where 1 = strongly disagree, 2 = somewhat disagree, 3 = neutral, 4 = somewhat agree, 5 = strongly agree:
Please choose the appropriate response for each item:

The tabular display increases my productivity.
The tabular display improves the quality of the work I do.
The tabular display enables me to accomplish tasks more quickly.
The tabular display supports critical aspects of my analysis.
The tabular display improves my job performance.
The tabular display enables me to explore relationships between different pieces of information.

34 For the TABULAR DISPLAY only, answer the following questions about ease of use where 1 = strongly disagree, 2 = somewhat disagree, 3 = neutral, 4 = somewhat agree, 5 = strongly agree:
Please choose the appropriate response for each item:

I would like to use the tabular display frequently.
I found the tabular display unnecessarily complex.
I thought the tabular display was easy to use.
I would need the support of a technical person to be able to use the tabular display.
I found that the various functions in the tabular display were well integrated.
I would imagine that most people would learn to use the tabular display easily.
I found the tabular display very cumbersome to use.
I felt confident using the tabular display.
I would need to learn a lot of things about the tabular display before I could get going with it.

35 For the PARALLEL COORDINATES DISPLAY only, answer the following questions about usefulness where 1 = strongly disagree, 2 = somewhat disagree, 3 = neutral, 4 = somewhat agree, 5 = strongly agree:

Please choose the appropriate response for each item:
The parallel coordinates display increases my productivity.
The parallel coordinates display improves the quality of the work I do.
The parallel coordinates display enables me to accomplish tasks more quickly.
The parallel coordinates display supports critical aspects of my analysis.
The parallel coordinates display improves my job performance.
The parallel coordinates display enables me to explore relationships between different pieces of information.

36 For the PARALLEL COORDINATES DISPLAY only, answer the following questions about ease of use where 1 = strongly disagree, 2 = somewhat disagree, 3 = neutral, 4 = somewhat agree, 5 = strongly agree:
Please choose the appropriate response for each item:

I would like to use the parallel coordinates display frequently.
I found the parallel coordinates display unnecessarily complex.
I thought the parallel coordinates display was easy to use.
I would need the support of a technical person to be able to use the parallel coordinates display.
I found that the various functions in the parallel coordinates display were well integrated.
I would imagine that most people would learn to use the parallel coordinates display easily.
I found the parallel coordinates display very cumbersome to use.
I felt confident using the parallel coordinates display.
I would need to learn a lot of things about the parallel coordinates display before I could get going with it.

37 For the NODE-LINK DISPLAY only, answer the following questions about usefulness where 1 = strongly disagree, 2 = somewhat disagree, 3 = neutral, 4 = somewhat agree, 5 = strongly agree:
Please choose the appropriate response for each item:
The node-link display increases my productivity.
The node-link display improves the quality of the work I do.
The node-link display enables me to accomplish tasks more quickly.
The node-link display supports critical aspects of my analysis.
The node-link display improves my job performance.
The node-link display enables me to explore relationships between different pieces of information.

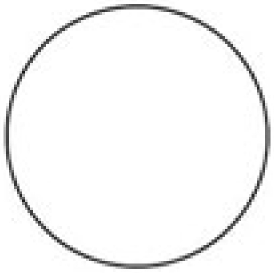
38 For the NODE-LINK DISPLAY only, answer the following questions about ease of use where 1 = strongly disagree, 2 = somewhat disagree, 3 = neutral, 4 = somewhat agree, 5 = strongly agree:
Please choose the appropriate response for each item:
I would like to use the node-link display frequently.
I found the node-link display unnecessarily complex.
I thought the node-link display was easy to use.
I would need the support of a technical person to be able to use the node-link display.
I found that the various functions in the node-link display were well integrated.
I would imagine that most people would learn to use the node-link display easily.

I found the node-link display very cumbersome to use.
I felt confident using the node-link display.
I would need to learn a lot of things about the node-link display before I could get going with it.

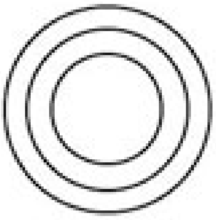
Post-Task Questionnaire, Analysis Survey #1

39 Answer the following questions based on the node representations provided.

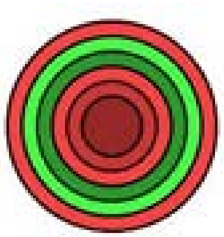
40 From the following set of figures, which node representation is best suited for representing the activity of the system (i.e., top talker)?



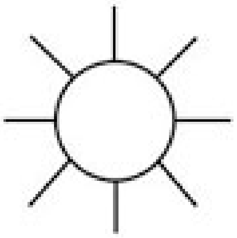
G



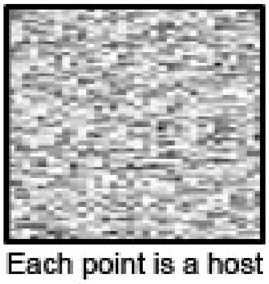
H



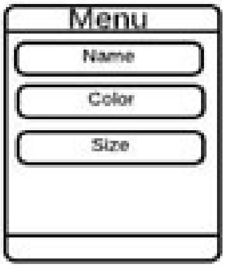
I



J



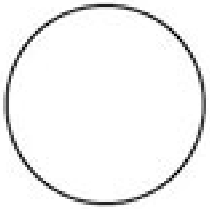
O



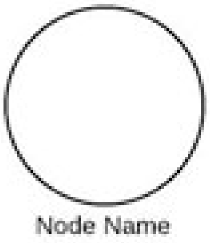
P

Please choose **only one** of the following:
[G, H, I, J, O, P]

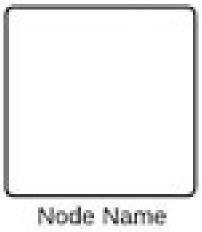
41 From the following set of figures, which node representation is best suited for labeling a system?



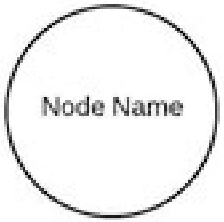
A



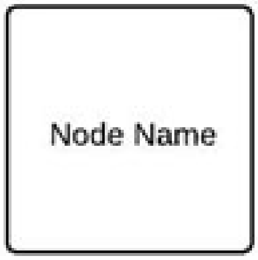
K



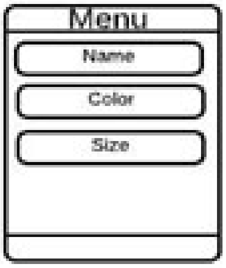
L



M



N

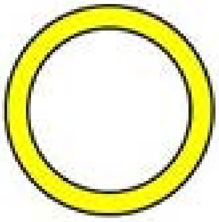


P

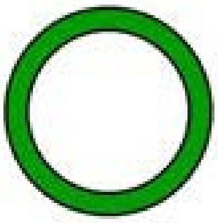
Please choose **only one** of the following:

[A, K, L, M, N, P]

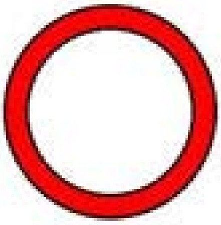
42 From the following set of figures, which node representation is best suited for representing the number of users located at a system?



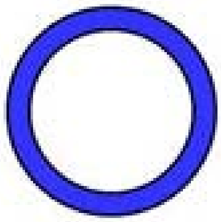
C



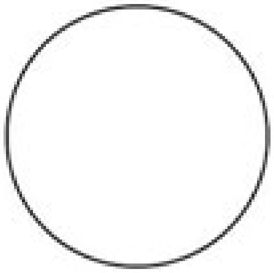
D



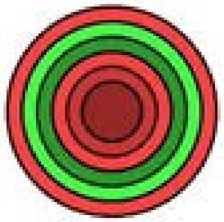
E



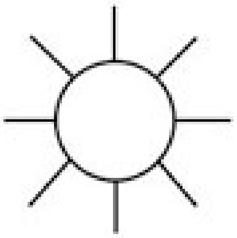
F



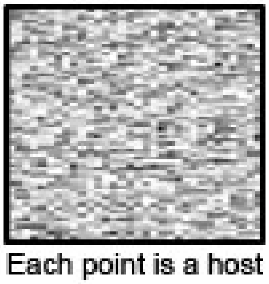
G



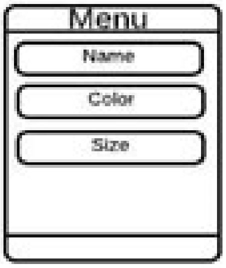
I



J



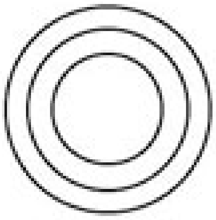
O



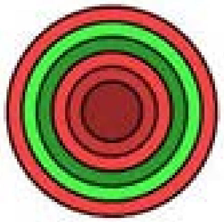
P

Please choose **only one** of the following:
[C, D, E, F, G, I, J, O, P]

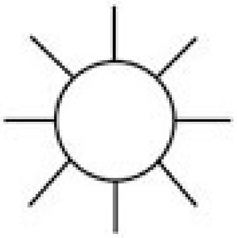
43 From the following set of figures, which node representation is best suited for representing the relevant past history of a system?



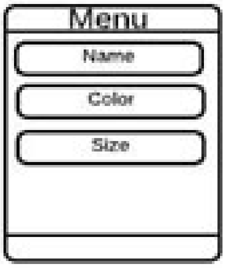
H



I



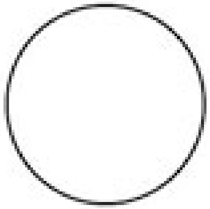
J



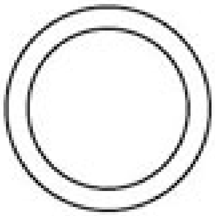
P

Please choose **only one** of the following:
[H, I, J, P]

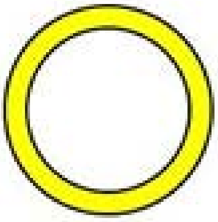
44 The next set of questions ask you to consider all of the following node representations.



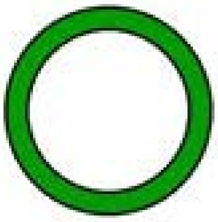
A



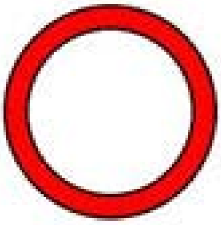
B



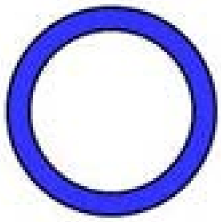
C



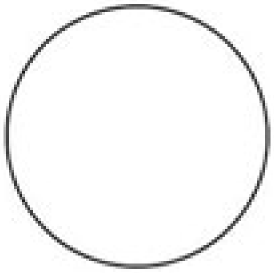
D



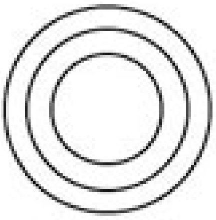
E



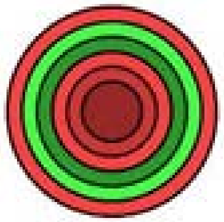
F



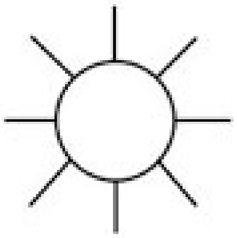
G



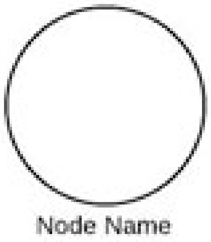
H



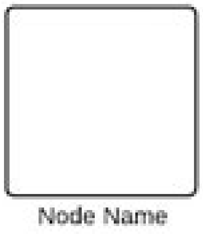
I



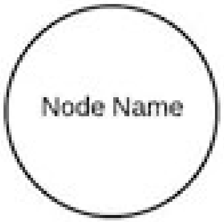
J



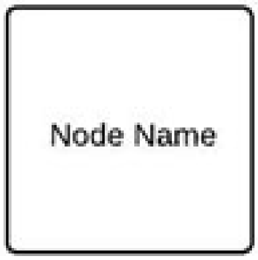
K



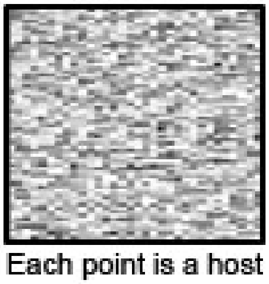
L



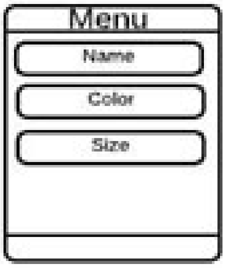
M



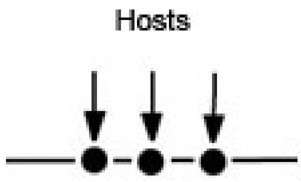
N



O



P



Q

45 Which node representation is best suited to represent an active system?
Please choose **only one** of the following:
[A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q]

46 Which node representation is best suited to represent an inactive system?
Please choose **only one** of the following:
[A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q]

47 Which node representation is best suited to represent a system under attack?
Please choose **only one** of the following:
[A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q]

48 Which node representation is best suited to represent a system that is vulnerable?
Please choose **only one** of the following:
[A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q]

49 Which node representation is best suited to represent a system that has been compromised?
Please choose **only one** of the following:
[A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q]

50 Which node representation is best suited to represent a high priority system?
Please choose **only one** of the following:
[A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q]

51 Which node representation is best suited to represent a low priority system?
Please choose **only one** of the following:
[A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q]

52 Rank the following network parameters in terms of relevance to analysis.

All your answers must be different.

Please number each box in order of preference from 1 to 14

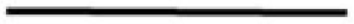
- CPU load
- Number of users
- Number of connections
- Network bandwidth usage
- Disk usage (%)
- Memory usage (%)
- Number of alerts generated

- Type of alerts generated
- Previous identification of issues with a specific system
- Median size of packets
- Connections asymmetry
- Operating system type
- System priority
- Other

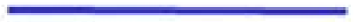
53 If you included "Other" in your ranking for the previous question, please specify the other parameter(s) of relevance to analysis.
Please write your answer here:

Post-Task Questionnaire, Analysis Survey #2

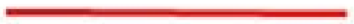
54 Answer the following questions based on the link representations provided.



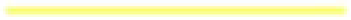
A



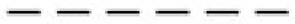
B



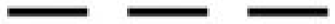
C



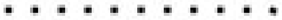
D



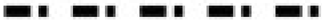
E



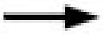
F



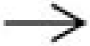
G



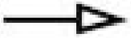
H



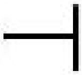
I



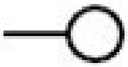
J



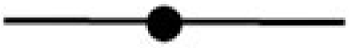
K



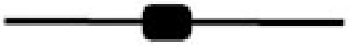
L



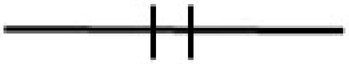
M



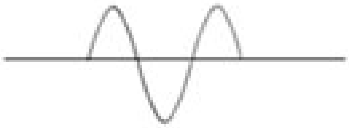
N



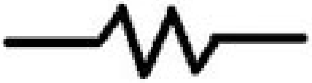
O



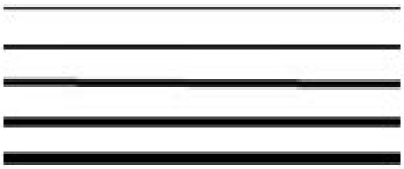
P



Q



R



(Line thickness is a parameter that represents load)

S

55 Which link representation is best suited to represent a connection from one system to another system?
Please choose **only one** of the following:
[A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S]

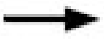
56 Which link representation is best suited to represent users with multiple connections?
Please choose **only one** of the following:
[A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S]

57 Which link representation is best suited to represent a TCP connection?
Please choose **only one** of the following:
[A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S]

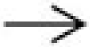
58 Which link representation is best suited to represent a UDP connection?
Please choose **only one** of the following:
[A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S]

59 Which link representation is best suited to represent access to a Network File System (NFS)?
Please choose **only one** of the following:
[A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S]

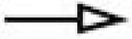
60 From the set I through M, which link representation is best suited to represent connections to a server?



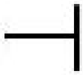
I



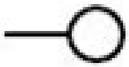
J



K



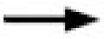
L



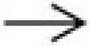
M

Please choose **only one** of the following:
[I, J, K, L, M]

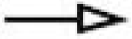
61 From the set I through M, which link representation is best suited to represent connections to a client?



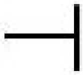
I



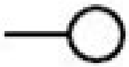
J



K



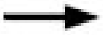
L



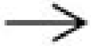
M

Please choose **only one** of the following:
[I, J, K, L, M]

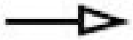
62 From the set I through M, which link representation is best suited to represent connections to a UNIX system?



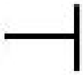
I



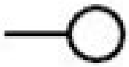
J



K



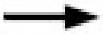
L



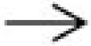
M

Please choose **only one** of the following:
[I, J, K, L, M]

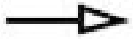
63 From the set I through M, which link representation is best suited to represent connections to a Windows system?



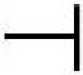
I



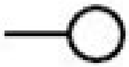
J



K



L



M

Please choose **only one** of the following:
[I, J, K, L, M]

Approved for public release; distribution is unlimited.

64 Considering all representations A-S, which link representation is best suited for representing CONUS connections?

Please choose **only one** of the following:

[A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S]

65 Which link representation is best suited for representing OCONUS connections?

Please choose **only one** of the following:

[A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S]

66 Which link representation is best suited for representing activity that generated an alert?

Please choose **only one** of the following:

[A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S]

67 Which link representation is best suited for representing the connection *from* a system under attack?

Please choose **only one** of the following:

[A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S]

68 Which link representation is best suited for representing the connection *to* a system under attack?

Please choose **only one** of the following:

[A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S]

69 Which link representation is best suited for representing an unauthorized system connection?

Please choose **only one** of the following:

[A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S]

70 Which link representation is best suited for representing normal traffic communications between systems?

Please choose **only one** of the following:

[A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S]

71 Which link representation is best suited for asymmetry of connections between inbound and outbound?

Please choose **only one** of the following:

[A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S]

72 Which link representation is best suited for representing the number of connections over the past 5 minutes?

Please choose **only one** of the following:

[A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S]

73 Which link representation is best suited for representing the number of connections over the past 1 hour?

Please choose **only one** of the following:

[A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S]

74 Which link representation is best suited for representing the number of connections over the past 24 hours?

Please choose **only one** of the following:

[A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S]

Post-Task Questionnaire, Subjective Survey #1

75 For the following questions, indicate with your opinion on a scale of 1 through 5, where 1 = strongly disagree, 2 = somewhat disagree, 3 = neutral, 4 = somewhat agree, 5 = strongly agree:

Please choose the appropriate response for each item:

The graphical displays were visually more appealing to the eye than the tabular display.

I easily understood the visualization of the ***graphical*** displays.

I easily understood the visualization of the ***tabular*** display.

The manipulation of the visualization's features of the ***graphical*** displays was easy.

The manipulation of the visualization's features of the ***tabular*** display was easy.

I was able to identify all of the intrusion alerts on the ***graphical*** displays.

I was able to identify all of the intrusion alerts on the ***tabular*** displays.

I was able to identify all of the network intrusions on the ***graphical*** displays.

I was able to identify all of the network intrusions on the ***tabular*** display.

The demo training provided by the investigators enabled me to use the visual displays effectively on the cyber analysis task.

I was able to complete my tasks better with the tabular display than the graphical displays.

I was able to complete my tasks better with the graphical displays than the tabular displays.

I prefer the graphical displays to the tabular display.

I prefer the tabular display to the graphical displays.

I recommend that the use of the graphical displays be incorporated into analyst's cyber-security systems.

I recommend that the use of the tabular display be incorporated into analyst's cyber-security systems.

I do not recommend that the graphical displays be incorporated into analyst's cyber-security systems

I do not recommend that the tabular display be incorporated into analyst's cyber-security systems.

Please choose the appropriate response for each item:

Post-Task Questionnaire, Subjective Survey #2

76 Overall, how would you rate the usefulness of the graphical layouts?

Please choose **only one** of the following:

[Excellent, Good, Fair, Poor]

77 Overall, how would you rate the usefulness of the tabular layout?
Please choose **only one** of the following:
[Excellent, Good, Fair, Poor]

78 Overall, how would you rate the appearance of the graphical layouts?
Please choose **only one** of the following:
[Excellent, Good, Fair, Poor]

79 Overall, how would you rate the appearance of the tabular layout?
Please choose **only one** of the following:
[Excellent, Good, Fair, Poor]

80 What components of the visual displays (table, parallel coordinates, and node link display) were most effective?
Please write your answer here:
[Input box]

81 What aspects of the visualizations (table, parallel coordinates, and node link display) did you like best?
Please write your answer here:
[Input box]

82 What aspects of the visualizations (table, parallel coordinates, and node link display) did you not like?
Please write your answer here:
[Input box]

83 What aspects of the visualizations (table, parallel coordinates, and node link display) helped you to identify intrusions?
Please write your answer here:
[Input box]

Post-Task Questionnaire, Analysis Survey #3

84 On a scale from 1 to 5 where 1 is "Very Sad", 3 is "Neutral", and 5 is "Very Happy", which of the following best describes your current emotional state?
Please choose **only one** of the following:
[1: Very Sad, 2: Sad, 3: Neutral, 4: Happy, 5: Very Happy]

85 Which is better at the following tasks, Machine or Human?
Please choose the appropriate response for each item:

Analyzing data [Machine, Human]
Detecting anomalies (where an anomaly is abnormal behavior on a security network) [Machine, Human]

86 Which type of display do you find most useful in analyzing security data?
Please choose **all** that apply:

[Tables, Textual, Line & Bar Graphs, Simple Graphs, Symbolic Shapes, Other (please specify)]

87 Please specify the other type of display you find useful in analyzing data.
Only answer this question if the following conditions are met:
Answer was at question '86 [3]' (Which type of display do you find most useful in analyzing security data?)
Please write your answer here:

[Input box]

88 What features of the visual displays were most useful for completing your tasks in this study?
Please write your answer here:

[Input box]

89 What did you learn from this study?
Please write your answer here:

[Input box]

90 In all three visual displays, what happened after your selections of intrusion attempts were submitted?
Please choose **only one** of the following:

[The task timer restarted, The submitted selections were grayed out and unable to be selected again, An alarm bell sounded]

91 On the "node-link" display, how could you view more information about an intrusion attempt?
Please choose **only one** of the following:

[Press the space bar and type in the node number of interest, Click and drag over the portion of the screen of interest, Hover your mouse over the marker of interest]

92 How effective do you believe these three different visual layouts will be in cyber analysis tasks?
Please choose the appropriate response for each item [1-5]:
Tabular
Parallel coordinates
Node-link

93 How much do you like these displays depending on their potential effectiveness of use?

Please choose the appropriate response for each item [1-Strongly dislike, 2-Somewhat dislike, 3-Neutral, 4-Somewhat like, 5-Strongly like]:

Tabular
Parallel coordinates
Node-link

Thank you for taking part in this study!

Submit your survey.
Thank you for completing this survey.

Appendix B. Display Pseudocode

This appendix appears in its original form, without editorial change.

This appendix presents pseudocode representing the display code (originally written in HTML and JavaScript). Note that code appearing in brackets (< >) refers to a user interface element.

B.1 Main Index

There is only a small amount of code embedded on the index.htm page. A script directs the user to the appropriate display, based on their assigned (randomized) sequence.

```
on <inputBox> submit then {
      call function openNextDisplay;
      }

function openNextDisplay {
      get url from array sequenceTable;
      set browserUrl = url;
}
```

B.2 Tabular

The tabular display shows an interactive table that reads data from a csv file on the server. The display requires the following JavaScript libraries:

- jQuery (http://jquery.com/)
- jQuery UI (http://jqueryui.com/)
- jQuery countdownTimer (http://plugins.jquery.com/countdownTimer/)
- dhtmlxGrid (http://dhtmlx.com/docs/products/dhtmlxGrid/)

```
on pageLoad then {
      call function initializeTable;
      initialize countdownTimer with value ’20 min’;
      get trueAnswers from array trueAnswersTable;
      get currentTime;
      set startTime = currentTime;
      set <progressBar> = 0;
      set <accuracyOutput> = ’70.5%’;
      set numberOfIdentifiedAlerts = 0;
      set totalAlerts = 139;
      set percentLimit = 50;
      set selectedRows = null;
      set savedRecords = null;
      set clickLogString = null;
      set submissionTimes = null;
}

function initializeTable {
      set headerColumns = ‘ID’,’Time’,’Src Entity’,’Src
Port’,’Dst Entity’,’Dst Port’,’Dst Country’,’Alert’;
      load allRows from ‘csvData.csv’;
      enable columnSorting;
      enable selectFiltering;
      enable rowHovering;
      initialize table with dhtmlxGrid;
}

on mouseClick then {
COMMENT: All clicks are logged for later analysis.

      get currentTime;
      get clickLocation;
      append currentTime and clickLocation to clickLogString;
}

on <submitIntrusionButton> click then {
      if selectedRows==null then {
            alert ‘You have not selected any data.’
      } otherwise then {
            show <confirmPrompt>;
      }
}

on <confirmButton> click then {
      get selectedRows;
      for each selectedRows {
            set rowColor = ‘gray’;
            lock selectedRow;
      }
      append selectedRows to savedRecords;
      set selectedRows = null;
      increase numberOfIdentifiedAlerts by number of
selectedRows;
      set percentComplete = numberOfIdentifiedAlerts/totalAlerts
* 100;
      set <progressBar> = percentComplete / percentLimit;
      if percentComplete > percentLimit then {
            call function endSession;
      }
      call function getAccuracy;
}

function endSession {
      get currentTime;
      set endTime = currentTime;
      append startTime and endTime to submissionTimes;
      save submissionTimes, clickLogString, and savedRecords to
file ‘userId#-table.txt’;
      call function openNextDisplay;
}

function getAccuracy {
      set truePositives = 0;
      set falsePositives = 0;
      set falseNegatives = 0;
      for each savedRecords {
            if savedRecord is in trueAnswers then {
                  increase truePositives by 1;
            } otherwise then {
                  increase falsePositives by 1;
            }
      }
      for each trueAnswers {
            if trueAnswer is not in savedRecords then {
                  increase falseNegatives by 1;
            }
      }
      set trueNegatives = totalAlerts –
(truePositives+falsePositives+falseNegatives);
      set theAccuracy = (truePositives +
trueNegatives)/totalAlerts * 100;
      set <accuracyOutput> = theAccuracy;
      if theAccuracy has increased then {
            set color of <accuracyOutput> = ‘green’;
      } otherwise then {
            set color of <accuracyOutput> = ‘red’;
      }
}

on <sessionCompleteButton> click then {
      confirm ‘Are you sure you want to end this session?’;
      if confirm==TRUE then {
            call function endSession;
      }
}

function timesUp(){
COMMENT: called when countdownTimer reaches 0 (from initial value
of 20 min)

      call function endSession;
}

function openNextDisplay {
      get url from array sequenceTable;
      set browserUrl = url;
}
```

B.3 Parallel Coordinates

The parallel coordinates display shows a graph of parallel coordinates with connecting lines that reads data from a csv file on the server. The user may select a range of traces (“brushes”) on any of the axes. The display requires the following JavaScript libraries:

- jQuery (http://jquery.com/)
- jQuery UI (http://jqueryui.com/)

- jQuery countdownTimer (http://plugins.jquery.com/countdownTimer/)
- D3.js (http://d3js.org/)
- D3.parcoords.js (http://syntagmatic.github.io/parallel-coordinates/)

```
on pageLoad then {
      call function initializePc;
      initialize countdownTimer with value ’20 min’;
      get trueAnswers from array trueAnswersTable;
      get currentTime;
      set startTime = currentTime;
      set <progressBar> = 0;
      set <accuracyOutput> = ’70.5%’;
      set numberOfIdentifiedAlerts = 0;
      set totalAlerts = 139;
      set percentLimit = 50;
      set selectedTraces = null;
      set savedRecords = null;
      set clickLogString = null;
      set submissionTimes = null;
}

function initializePc {
      load allTraces from ‘csvData.csv’;
      enable brushing;
      enable shadows;
      set transparency = ‘40%’;
      set lineColor = ‘blue’;
      initialize graph with d3.parcoords;
}

function getColor(trace) {
COMMENT: When the data is loaded (or reloaded) into the parallel
coordinates display, the color for each trace is determined by
this function based on whether or not it is one of the saved
traces.

      if trace is in savedRecords then {
            return ‘gray’;
      } otherwise {
            return ‘blue’;
      }
}

on mouseClick then {
COMMENT: All clicks are logged for later analysis.

      get currentTime;
      get clickLocation;
      append currentTime and clickLocation to clickLogString;
}

on <submitIntrusionButton> click then {
      if selectedRows==null then {
            alert ‘You have not selected any data.’
      } otherwise then {
            show <confirmPrompt>;
      }
}

on <confirmButton> click then {
      get selectedTraces;
      append selectedTraces to savedRecords;
      increase numberOfIdentifiedAlerts by number of
selectedTraces;
      set percentComplete = numberOfIdentifiedAlerts/totalAlerts
* 100;
      set <progressBar> = percentComplete / percentLimit;
      if percentComplete > percentLimit then {
            call function endSession;
      }
      call function getAccuracy;
      reset brushes;
      for each trace in allTraces {
            call function getColor(trace);
            set color;
      }
}

function endSession {
      get currentTime;
      set endTime = currentTime;
      append startTime and endTime to submissionTimes;
      save submissionTimes, clickLogString, and savedRecords to
file ‘userId#-pc.txt’;
      call function openNextDisplay;
}

function getAccuracy {
      set truePositives = 0;
      set falsePositives = 0;
      set falseNegatives = 0;
      for each savedRecords {
            if savedRecord is in trueAnswers then {
                  increase truePositives by 1;
            } otherwise then {
                  increase falsePositives by 1;
            }
      }
      for each trueAnswers {
            if trueAnswer is not in savedRecords then {
                  increase falseNegatives by 1;
            }
      }
      set trueNegatives = totalAlerts –
(truePositives+falsePositives+falseNegatives);
      set theAccuracy = (truePositives +
trueNegatives)/totalAlerts * 100;
      set <accuracyOutput> = theAccuracy;
      if theAccuracy has increased then {
            set color of <accuracyOutput> = ‘green’;
      } otherwise then {
            set color of <accuracyOutput> = ‘red’;
      }
}

on <sessionCompleteButton> click then {
      confirm ‘Are you sure you want to end this session?’;
      if confirm==TRUE then {
            call function endSession;
      }
}

function timesUp(){
COMMENT: This function is called when countdownTimer reaches 0
(from initial value of 20 min).

      call function endSession;
}

function openNextDisplay {
      get url from array sequenceTable;
      set browserUrl = url;
}
```

B.4 Node-Link

The node-link display shows a graph of source and destination nodes with lines connecting them. The graph is a static image, and the code enhances the image by making it interactive (e.g., clickable markers). The markers are generated by using the “canvas” element of HTML 5. The display requires the following JavaScript libraries:

- jQuery (http://jquery.com/)
- jQuery UI (http://jqueryui.com/)
- jQuery countdownTimer (http://plugins.jquery.com/countdownTimer/)

```
on pageLoad then {
      call function initializeTooltip;
      initialize countdownTimer with value ’20 min’;
      get trueAnswers from array trueAnswersTable;
      get currentTime;
      set startTime = currentTime;
      set <progressBar> = 0;
      set <accuracyOutput> = ’70.5%’;
      set numberOfIdentifiedAlerts = 0;
      set totalAlerts = 139;
      set percentLimit = 50;
      set savedRecords = null;
      set clickLogString = null;
      set submissionTimes = null;
      set selectedMarkerArray = null;
      set deactivatedMarkerArray = null;
      set selectedCoordinateArray = null;
      set deactivatedCoordinateArray = null;
}

function initializeTooltip {
COMMENT: Each marker has an associated tooltip with alert details
that is coded into the html on the page. This function loads that
data into tooltips that appear when the user hovers over each
marker.

      enable mouseTracking;
      load tooltipContents from html;
}

on commaKeyPress {
COMMENT: This is a hack that clears all tooltips when the user
presses the comma key.

      clear tooltip;
}

on mouseClick then {
COMMENT: All clicks are logged for later analysis.

      get currentTime;
      get clickLocation;
      append currentTime and clickLocation to clickLogString;
}


on <submitIntrusionButton> click then {
      if selectedRows==null then {
            alert ‘You have not selected any data.’
      } otherwise then {
            show <confirmPrompt>;
      }
}

on <confirmButton> click then {
      append selectedMarkerArray to savedRecords;
      increase numberOfIdentifiedAlerts by number of
selectedMarkerArray;
      set percentComplete = numberOfIdentifiedAlerts/totalAlerts
* 100;
      set <progressBar> = percentComplete / percentLimit;
      if percentComplete > percentLimit then {
            call function endSession;
      }
      call function getAccuracy;
      call function deactivateMarkers;
}

function deactivateMarkers {
      for each selectedCoordinateArray {
            clear marker at (selectedCoordinateArray.X,
selectedCoordinateArray.Y);
            draw marker using color ‘gray’ at
(selectedCoordinateArray.X, selectedCoordinateArray.Y);
            append selectedCoordinateArray to
deactivatedCoordinateArray;
            append selectedMarkerArray to deactivatedIdArray;
      }
      set selectedCoordinateArray = null;
}

function selectMarker(x,y,r) {
COMMENT: Each marker has a link that is coded on the html page.
When the link is clicked, this function is called and parameters
x,y,r are passed to this function. This function toggles the red
active marker.

      if (x,y) is not in deactivatedCoordinateArray or
selectedCoordinateArray then {
            append (x,y) to selectedCoordinateArray;
            draw circle of radius r at (x,y) with color ‘red’;
            append selectedMarker to selectedMarkerArray;
      } otherwise if (x,y) is not in deactivatedCoordinateArray
and is in selectedCoordinateArray then {
            remove (x,y) from selectedCoordinateArray;
            clear circle at (x,y);
            remove selectedMarker from selectedMarkerArray;
      }
}

function endSession {
      get currentTime;
      set endTime = currentTime;
      append startTime and endTime to submissionTimes;
      save submissionTimes, clickLogString, and savedRecords to
file ‘userId#-nodelink.txt’;
      call function openNextDisplay;
}


function getAccuracy {
      set truePositives = 0;
      set falsePositives = 0;
      set falseNegatives = 0;
      for each savedRecords {
            if savedRecord is in trueAnswers then {
                  increase truePositives by 1;
            } otherwise then {
                  increase falsePositives by 1;
            }
      }
      for each trueAnswers {
            if trueAnswer is not in savedRecords then {
                  increase falseNegatives by 1;
            }
      }
      set trueNegatives = totalAlerts –
(truePositives+falsePositives+falseNegatives);
      set theAccuracy = (truePositives +
trueNegatives)/totalAlerts * 100;
      set <accuracyOutput> = theAccuracy;
      if theAccuracy has increased then {
            set color of <accuracyOutput> = ‘green’;
      } otherwise then {
            set color of <accuracyOutput> = ‘red’;
      }
}

on <sessionCompleteButton> click then {
      confirm ‘Are you sure you want to end this session?’;
      if confirm==TRUE then {
            call function endSession;
      }
}

function timesUp(){
COMMENT: called when countdownTimer reaches 0 (from initial value
of 20 min)

      call function endSession;
}

function openNextDisplay {
      get url from array sequenceTable;
      set browserUrl = url;
}
```

Appendix C. R Code for Calculating Objective Results from Data

This appendix appears in its original form, without editorial change.

Approved for public release; distribution is unlimited.

The code provided in this appendix may be used to calculate and summarize the various metrics for determining objective performance (TP, FP, accuracy, etc.). Directory of data files and participant ID's to use in the calculations must be defined at the beginning of the code. Several functions are provided that perform various calculations; to get a summary of results run `getSummarySDTAgg()`.

```
# This code assumes that your working directory has a sub-
 directory defined below that contains all data files, organized
 into folders with the participant ID's. The "true_answers.txt"
 file must also appear in the sub-directory.
sub_directory = "All/";
require(rjson);
# for calculations involving aggregate performance...
idArray =
 c('101M','112','128','146','175','261','274','298','333','340','4
 11','481','493','515','597','674','678','734','747','817','840','
 874','913','921'); #ARL participant ID’s


getResults <- function(id){
#id is the participant id
#example usage: getResults('101')$'tabular'

concat_file_ID = paste(sub_directory,id,"/id",id,"-
 table.txt",sep="");
tabular = fromJSON(file=concat_file_ID);
concat_file_ID = paste(sub_directory,id,"/id",id,"-
 pc.txt",sep="");
pc = fromJSON(file=concat_file_ID);
concat_file_ID = paste(sub_directory,id,"/id",id,"-
 nodelink.txt",sep="");
nodelink = fromJSON(file=concat_file_ID);
result<-list("tabular"=tabular,"pc"=pc,"nodelink"=nodelink);
return(result);
}

getTrueAnswers <- function(){
#can append $'easy', $'moderate', or $'hard' when calling this
 function
#example usage: getTrueAnswers()$'easy'

concat_file_ID = paste(sub_directory,"true_answers.txt",sep="");
return(fromJSON(file=concat_file_ID));

}

stripComments <- function(orig){
#strips comments associated with each set of responses and
 returns list with row ID's only
origM=NULL;
for (j in 1:length(orig)){
 k=1;
      for (i in 1:length(orig[[j]])){
           if (nchar(orig[[j]][i])<=3){ #alert id numbers are three
```

Approved for public release; distribution is unlimited.

```
 characters or less
             origM=cbind(origM,orig[[j]][i]);
             k=k+1;
         }
     }
 }
 return(origM[origM!=""]);
}

findResponseSDTAgg <- function(){
#finds aggregate response for all participants
#alertSet is set of answers to use for computing aggregate
 performance
#tip: to include all alert sets, use
 c(getTrueAnswers()$'easy',getTrueAnswers()$'moderate',getTrueAnsw
 ers()$'hard');

alertSet =
 c(getTrueAnswers()$'easy',getTrueAnswers()$'moderate',getTrueAnsw
 ers()$'hard');

TP_tabular = "";
TP_pc = "";
TP_nodelink = "";
TP_percent_tabular = "";
TP_percent_pc = "";
TP_percent_nodelink = "";
FP_tabular = "";
FP_pc = "";
FP_nodelink = "";
DUR_tabular = "";
DUR_pc = "";
DUR_nodelink = "";
ACC_tabular = "";
ACC_pc = "";
ACC_nodelink = "";
PREC_tabular = "";
PREC_pc = "";
PREC_nodelink = "";
F_tabular = "";
F_pc = "";
F_nodelink = "";
R_tabular = "";
R_pc = "";
R_nodelink = "";

for (idN in idArray){
TP_tabular =
 c(TP_tabular,findResponseSDT(getResults(idN)$'tabular',alertSet,1
 39)$'TP');
TP_pc =
 c(TP_pc,findResponseSDT(getResults(idN)$'pc',alertSet,139)$'TP');
TP_nodelink =
 c(TP_nodelink,findResponseSDT(getResults(idN)$'nodelink',alertSet
 ,139)$'TP');
TP_percent_tabular =
 c(TP_percent_tabular,findResponseSDT(getResults(idN)$'tabular',al
```

Approved for public release; distribution is unlimited.

```
 ertSet,139)$'TP_percent');
TP_percent_pc =
 c(TP_percent_pc,findResponseSDT(getResults(idN)$'pc',alertSet,139
 )$'TP_percent');
TP_percent_nodelink =
 c(TP_percent_nodelink,findResponseSDT(getResults(idN)$'nodelink',
 alertSet,139)$'TP_percent');
FP_tabular =
 c(FP_tabular,findResponseSDT(getResults(idN)$'tabular',alertSet,1
 39)$'FP');
FP_pc =
 c(FP_pc,findResponseSDT(getResults(idN)$'pc',alertSet,139)$'FP');
FP_nodelink =
 c(FP_nodelink,findResponseSDT(getResults(idN)$'nodelink',alertSet
 ,139)$'FP');
DUR_tabular = c(DUR_tabular,getTime(idN,'t')$'duration');
DUR_pc = c(DUR_pc,getTime(idN,'p')$'duration');
DUR_nodelink = c(DUR_nodelink,getTime(idN,'n')$'duration');
ACC_tabular =
 c(ACC_tabular,findResponseSDT(getResults(idN)$'tabular',alertSet,
 139)$'Accuracy');
ACC_pc =
 c(ACC_pc,findResponseSDT(getResults(idN)$'pc',alertSet,139)$'Accu
 racy');
ACC_nodelink =
 c(ACC_nodelink,findResponseSDT(getResults(idN)$'nodelink',alertSe
 t,139)$'Accuracy');
PREC_tabular =
 c(PREC_tabular,findResponseSDT(getResults(idN)$'tabular',alertSet
 ,139)$'Precision');
PREC_pc =
 c(PREC_pc,findResponseSDT(getResults(idN)$'pc',alertSet,139)$'Pre
 cision');
PREC_nodelink =
 c(PREC_nodelink,findResponseSDT(getResults(idN)$'nodelink',alertS
 et,139)$'Precision');
F_tabular =
 c(F_tabular,findResponseSDT(getResults(idN)$'tabular',alertSet,13
 9)$'Fscore');
F_pc =
 c(F_pc,findResponseSDT(getResults(idN)$'pc',alertSet,139)$'Fscore
 ');
F_nodelink =
 c(F_nodelink,findResponseSDT(getResults(idN)$'nodelink',alertSet,
 139)$'Fscore');
R_tabular =
 c(R_tabular,findResponseSDT(getResults(idN)$'tabular',alertSet,13
 9)$'Recall');
R_pc =
 c(R_pc,findResponseSDT(getResults(idN)$'pc',alertSet,139)$'Recall
 ');
R_nodelink =
 c(R_nodelink,findResponseSDT(getResults(idN)$'nodelink',alertSet,
 139)$'Recall');

}
```

```
result<-
 list("TP_tabular"=TP_tabular,"TP_pc"=TP_pc,"TP_nodelink"=TP_nodel
 ink,"TP_percent_tabular"=TP_percent_tabular,"TP_percent_pc"=TP_pe
 rcent_pc,"TP_percent_nodelink"=TP_percent_nodelink,"FP_tabular"=F
 P_tabular,"FP_pc"=FP_pc,"FP_nodelink"=FP_nodelink,"DUR_tabular"=D
 UR_tabular,"DUR_pc"=DUR_pc,"DUR_nodelink"=DUR_nodelink,"ACC_tabul
 ar"=ACC_tabular,"ACC_pc"=ACC_pc,"ACC_nodelink"=ACC_nodelink,"PREC
 _tabular"=PREC_tabular,"PREC_pc"=PREC_pc,"PREC_nodelink"=PREC_nod
 elink,"F_tabular"=F_tabular,"F_pc"=F_pc,"F_nodelink"=F_nodelink,"
 R_tabular"=R_tabular,"R_pc"=R_pc,"R_nodelink"=R_nodelink);

return(result);
}


findResponseSDT <- function(responseSet,answerSet,totalAlerts){
#function returns number of true positives TP, false positives
 FP, true negatives TN, and false negatives FN according to signal
 detection theory (SDT)
#responseSet is the vector of responses and answerSet is the
 vector of true answers, and totalAlerts is the number of all
 alerts in experiment
#responses returned by this function are divided into true
 positives TP, false positives FP, true negatives TN, and false
 negatives FN
#example usage:
 findResponseSDT(getResults('101')$'tabular',getTrueAnswers()$'har
 d',139)$'TP'


if (length(responseSet)>0){

responseSetM = stripComments(responseSet);
completeAnswerSet =
 c(getTrueAnswers()$'easy',getTrueAnswers()$'moderate',getTrueAnsw
 ers()$'hard'); # get all true answers, since this is how we have
 to calculate FP


# TP calculations


 answerSet_TP = answerSet; # we expect only a subset of answers,
 i.e., easy, moderate, or hard

 for (j in responseSetM){
     for (i in answerSet){
          if (i %in% j){
               answerSet_TP = answerSet_TP[which(answerSet_TP!=i)];
 #remove this answer so it's not counted twice
          }
     }
 }

 TP = length(answerSet)-length(answerSet_TP);

 if (length(answerSet_TP)==0) { TP_percent = 100;}
 else{ tp = 100*(length(answerSet)-
 length(answerSet_TP))/length(answerSet);
     TP_percent = format(round(tp,1),nsmall=1);}

# TP calculations (all alert sets)


 answerSet_TP = completeAnswerSet;

 for (j in responseSetM){
     for (i in completeAnswerSet){
         if (i %in% j){
             answerSet_TP = answerSet_TP[which(answerSet_TP!=i)];
 #remove this answer so it's not counted twice
         }
     }
 }

 TP_all = length(completeAnswerSet)-length(answerSet_TP);

 if (length(answerSet_TP)==0) { TP_percent_all = 100;}
 else{ tp = 100*(length(completeAnswerSet)-
 length(answerSet_TP))/length(completeAnswerSet);
     TP_percent_all = format(round(tp,1),nsmall=1);}


# FP calculations (all alert sets)

 FP = abs(length(responseSetM)-TP_all);

# FN calculations (all alert sets)

 FN = abs(length(completeAnswerSet)-TP_all);

# TN calculations (all alert sets)

 TN = totalAlerts - (TP_all+FP+FN);

# See http://en.wikipedia.org/wiki/Sensitivity_and_specificity
 for metrics below

# Accuracy

 Accuracy = (TP_all+TN)/totalAlerts;
 Accuracy = format(round(Accuracy,2),nsmall=2);


# Sensitivity (True Positive Rate) / Recall

 Sensitivity = TP_all/(TP_all+FN);
 Sensitivity = format(round(Sensitivity,2),nsmall=2);

# Specificity (True Negative Rate)

 Specificity = TN/(TN+FP);
 Specificity = format(round(Specificity,2),nsmall=2);
```

```
# Precision (Positive Predictive Value)

 Precision = TP_all/(TP_all+FP);
 Precision = format(round(Precision,2),nsmall=2);

# Recall

 Recall = TP_all/(TP_all+FN);
 Recall = format(round(Recall,2),nsmall=2);

# Fscore
# = 2 * (Precision*Recall)/(Precision+Recall);

 Fscore = 2*TP_all/(2*TP_all+FP+FN);
 Fscore = format(round(Fscore,2),nsmall=2);


result<-
 list("TP"=TP,"FP"=FP,"FN"=FN,"TN"=TN,"TP_percent"=TP_percent,"TP_
 all"=TP_all,"TP_percent_all"=TP_percent_all,"Sensitivity"=Sensiti
 vity,"Specificity"=Specificity,"Precision"=Precision,"Fscore"=Fsc
 ore,"Accuracy"=Accuracy,"Recall"=Recall);
return(result);

} else {
result<-
 list("TP"='NA',"FP"='NA',"FN"='NA',"TN"='NA',"TP_percent"='NA',"T
 P_all"='NA',"TP_percent_all"='NA',"Sensitivity"='NA',"Specificity
 "='NA',"Precision"='NA',"Fscore"='NA',"Accuracy"='NA');
}

return(result);

}

getSummarySDT <- function(id){
#for a given participant id, returns summary of TP/FP/TN/FN for
 tabular, pc, and node-link displays using the easy, moderate, and
 hard alert sets
#example usage: getSummarySDT('128');

totalAlerts = 139;
cat('Results for participant ID: ',id,'\n',sep='');
cat('--EASY ALERT SET--\n');
cat('\t\tTab\t\tPC\t\tNL\n');
cat('TP',findResponseSDT(getResults(id)$'tabular',getTrueAnswers(
 )$'easy',totalAlerts)$'TP',findResponseSDT(getResults(id)$'pc',ge
 tTrueAnswers()$'easy',totalAlerts)$'TP',findResponseSDT(getResult
 s(id)$'nodelink',getTrueAnswers()$'easy',totalAlerts)$'TP','\n',s
 ep="\t\t");
cat('TP%',findResponseSDT(getResults(id)$'tabular',getTrueAnswers
 ()$'easy',totalAlerts)$'TP_percent',findResponseSDT(getResults(id
 )$'pc',getTrueAnswers()$'easy',totalAlerts)$'TP_percent',findResp
 onseSDT(getResults(id)$'nodelink',getTrueAnswers()$'easy',totalAl
 erts)$'TP_percent','\n',sep="\t\t");
cat('--MODERATE ALERT SET--\n');
cat('\t\tTab\t\tPC\t\tNL\n');
cat('TP',findResponseSDT(getResults(id)$'tabular',getTrueAnswers(
 )$'moderate',totalAlerts)$'TP',findResponseSDT(getResults(id)$'pc
 ',getTrueAnswers()$'moderate',totalAlerts)$'TP',findResponseSDT(g
 etResults(id)$'nodelink',getTrueAnswers()$'moderate',totalAlerts)
 $'TP','\n',sep="\t\t");
cat('TP%',findResponseSDT(getResults(id)$'tabular',getTrueAnswers
 ()$'moderate',totalAlerts)$'TP_percent',findResponseSDT(getResult
 s(id)$'pc',getTrueAnswers()$'moderate',totalAlerts)$'TP_percent',
 findResponseSDT(getResults(id)$'nodelink',getTrueAnswers()$'moder
 ate',totalAlerts)$'TP_percent','\n',sep="\t\t");
cat('--HARD ALERT SET--\n');
cat('\t\tTab\t\tPC\t\tNL\n');
cat('TP',findResponseSDT(getResults(id)$'tabular',getTrueAnswers(
 )$'hard',totalAlerts)$'TP',findResponseSDT(getResults(id)$'pc',ge
 tTrueAnswers()$'hard',totalAlerts)$'TP',findResponseSDT(getResult
 s(id)$'nodelink',getTrueAnswers()$'hard',totalAlerts)$'TP','\n',s
 ep="\t\t");
cat('TP%',findResponseSDT(getResults(id)$'tabular',getTrueAnswers
 ()$'hard',totalAlerts)$'TP_percent',findResponseSDT(getResults(id
 )$'pc',getTrueAnswers()$'hard',totalAlerts)$'TP_percent',findResp
 onseSDT(getResults(id)$'nodelink',getTrueAnswers()$'hard',totalAl
 erts)$'TP_percent','\n',sep="\t\t");
cat('--ALL ALERTS--\n');
cat('\t\tTab\t\tPC\t\tNL\n');
cat('TP',findResponseSDT(getResults(id)$'tabular',getTrueAnswers(
 ),totalAlerts)$'TP_all',findResponseSDT(getResults(id)$'pc',getTr
 ueAnswers(),totalAlerts)$'TP_all',findResponseSDT(getResults(id)$
 'nodelink',getTrueAnswers(),totalAlerts)$'TP_all','\n',sep="\t\t"
 );
cat('TP%',findResponseSDT(getResults(id)$'tabular',getTrueAnswers
 (),totalAlerts)$'TP_percent_all',findResponseSDT(getResults(id)$'
 pc',getTrueAnswers(),totalAlerts)$'TP_percent_all',findResponseSD
 T(getResults(id)$'nodelink',getTrueAnswers(),totalAlerts)$'TP_per
 cent_all','\n',sep="\t\t");
cat('FP',findResponseSDT(getResults(id)$'tabular',getTrueAnswers(
 ),totalAlerts)$'FP',findResponseSDT(getResults(id)$'pc',getTrueAn
 swers(),totalAlerts)$'FP',findResponseSDT(getResults(id)$'nodelin
 k',getTrueAnswers(),totalAlerts)$'FP','\n',sep="\t\t");
cat('FN',findResponseSDT(getResults(id)$'tabular',getTrueAnswers(
 ),totalAlerts)$'FN',findResponseSDT(getResults(id)$'pc',getTrueAn
 swers(),totalAlerts)$'FN',findResponseSDT(getResults(id)$'nodelin
 k',getTrueAnswers(),totalAlerts)$'FN','\n',sep="\t\t");
cat('TN',findResponseSDT(getResults(id)$'tabular',getTrueAnswers(
 ),totalAlerts)$'TN',findResponseSDT(getResults(id)$'pc',getTrueAn
 swers(),totalAlerts)$'TN',findResponseSDT(getResults(id)$'nodelin
 k',getTrueAnswers(),totalAlerts)$'TN','\n',sep="\t\t");
cat('--OTHER METRICS--\n');
cat('\t\tTab\t\tPC\t\tNL\n');
cat('DUR-
 M',getTime(id,'t')$'duration',getTime(id,'p')$'duration',getTime(
 id,'n')$'duration','\n',sep="\t\t");
cat('ACC',findResponseSDT(getResults(id)$'tabular',getTrueAnswers
 (),totalAlerts)$'Accuracy',findResponseSDT(getResults(id)$'pc',ge
 tTrueAnswers(),totalAlerts)$'Accuracy',findResponseSDT(getResults
 (id)$'nodelink',getTrueAnswers(),totalAlerts)$'Accuracy','\n',sep
 ="\t\t");
cat('PREC',findResponseSDT(getResults(id)$'tabular',getTrueAnswer
 s(),totalAlerts)$'Precision',findResponseSDT(getResults(id)$'pc',
 getTrueAnswers(),totalAlerts)$'Precision',findResponseSDT(getResu
 lts(id)$'nodelink',getTrueAnswers(),totalAlerts)$'Precision','\n'
 ,sep="\t\t");
cat('SENS',findResponseSDT(getResults(id)$'tabular',getTrueAnswer
 s(),totalAlerts)$'Sensitivity',findResponseSDT(getResults(id)$'pc
 ',getTrueAnswers(),totalAlerts)$'Sensitivity',findResponseSDT(get
 Results(id)$'nodelink',getTrueAnswers(),totalAlerts)$'Sensitivity
 ','\n',sep="\t\t");
cat('SPEC',findResponseSDT(getResults(id)$'tabular',getTrueAnswer
 s(),totalAlerts)$'Specificity',findResponseSDT(getResults(id)$'pc
 ',getTrueAnswers(),totalAlerts)$'Specificity',findResponseSDT(get
 Results(id)$'nodelink',getTrueAnswers(),totalAlerts)$'Specificity
 ','\n',sep="\t\t");
cat('FSC',findResponseSDT(getResults(id)$'tabular',getTrueAnswers
 (),totalAlerts)$'Fscore',findResponseSDT(getResults(id)$'pc',getT
 rueAnswers(),totalAlerts)$'Fscore',findResponseSDT(getResults(id)
 $'nodelink',getTrueAnswers(),totalAlerts)$'Fscore','\n',sep="\t\t
 ");
}

getSummarySDTAgg <- function(){
#returns aggregate summary of TP/FP/TN/FN for tabular, pc, and
 node-link displays FOR ALL PARTICIPANTS identified in idArray at
 top
#note this takes 5-10 minutes to calculate

cat('Calculating...this may take a few minutes.');

# get arrays of responses from findResponseSDTAgg() and do it
 only once to save computation time
TP_tabular <<- findResponseSDTAgg()$'TP_tabular';
TP_pc <<- findResponseSDTAgg()$'TP_pc';
TP_nodelink <<- findResponseSDTAgg()$'TP_nodelink';
TP_percent_tabular <<- findResponseSDTAgg()$'TP_percent_tabular';
TP_percent_pc <<- findResponseSDTAgg()$'TP_percent_pc';
TP_percent_nodelink <<-
 findResponseSDTAgg()$'TP_percent_nodelink';
FP_tabular <<- findResponseSDTAgg()$'FP_tabular';
FP_pc <<- findResponseSDTAgg()$'FP_pc';
FP_nodelink <<- findResponseSDTAgg()$'FP_nodelink';
DUR_tabular <<- findResponseSDTAgg()$'DUR_tabular';
DUR_pc <<- findResponseSDTAgg()$'DUR_pc';
DUR_nodelink <<- findResponseSDTAgg()$'DUR_nodelink';
ACC_tabular <<- findResponseSDTAgg()$'ACC_tabular';
ACC_pc <<- findResponseSDTAgg()$'ACC_pc';
ACC_nodelink <<- findResponseSDTAgg()$'ACC_nodelink';
PREC_tabular <<- findResponseSDTAgg()$'PREC_tabular';
PREC_pc <<- findResponseSDTAgg()$'PREC_pc';
PREC_nodelink <<- findResponseSDTAgg()$'PREC_nodelink';
F_tabular <<- findResponseSDTAgg()$'F_tabular';
F_pc <<- findResponseSDTAgg()$'F_pc';
F_nodelink <<- findResponseSDTAgg()$'F_nodelink';
R_tabular <<- findResponseSDTAgg()$'R_tabular';
R_pc <<- findResponseSDTAgg()$'R_pc';
R_nodelink <<- findResponseSDTAgg()$'R_nodelink';
```

```
cat('Results for participant IDs: ','\n',sep='');
cat(idArray,'\n',sep=",");
cat('--ALL ALERTS--\n');
cat('\t\tTab\t\tPC\t\tNL\n');
cat('n',sum(!is.na(as.numeric(TP_tabular))),sum(!is.na(as.numeric
 (TP_pc))),sum(!is.na(as.numeric(TP_nodelink))),'\n',sep="\t\t");
cat('TP-
 mean',mean(as.numeric(TP_tabular),na.rm=TRUE),mean(as.numeric(TP_
 pc),na.rm=TRUE),mean(as.numeric(TP_nodelink),na.rm=TRUE),'\n',sep
 ="\t\t");
cat('TP-
 sd',sd(as.numeric(TP_tabular),na.rm=TRUE),sd(as.numeric(TP_pc),na
 .rm=TRUE),sd(as.numeric(TP_nodelink),na.rm=TRUE),'\n',sep="\t\t")
 ;
cat('TP%-
 mean',mean(as.numeric(TP_percent_tabular),na.rm=TRUE),mean(as.num
 eric(TP_percent_pc),na.rm=TRUE),mean(as.numeric(TP_percent_nodeli
 nk),na.rm=TRUE),'\n',sep="\t\t");
cat('TP%-
 sd',sd(as.numeric(TP_percent_tabular),na.rm=TRUE),sd(as.numeric(T
 P_percent_pc),na.rm=TRUE),sd(as.numeric(TP_percent_nodelink),na.r
 m=TRUE),'\n',sep="\t\t");
cat('FP-
 mean',mean(as.numeric(FP_tabular),na.rm=TRUE),mean(as.numeric(FP_
 pc),na.rm=TRUE),mean(as.numeric(FP_nodelink),na.rm=TRUE),'\n',sep
 ="\t\t");
cat('FP-
 sd',sd(as.numeric(FP_tabular),na.rm=TRUE),sd(as.numeric(FP_pc),na
 .rm=TRUE),sd(as.numeric(FP_nodelink),na.rm=TRUE),'\n',sep="\t\t")
 ;
cat('DUR-
 mean',mean(as.numeric(DUR_tabular),na.rm=TRUE),mean(as.numeric(DU
 R_pc),na.rm=TRUE),mean(as.numeric(DUR_nodelink),na.rm=TRUE),'\n',
 sep="\t\t");
cat('DUR-
 sd',sd(as.numeric(DUR_tabular),na.rm=TRUE),sd(as.numeric(DUR_pc),
 na.rm=TRUE),sd(as.numeric(DUR_nodelink),na.rm=TRUE),'\n',sep="\t\
 t");
cat('ACC-
 mean',mean(as.numeric(ACC_tabular),na.rm=TRUE),mean(as.numeric(AC
 C_pc),na.rm=TRUE),mean(as.numeric(ACC_nodelink),na.rm=TRUE),'\n',
 sep="\t\t");
cat('ACC-
 sd',sd(as.numeric(ACC_tabular),na.rm=TRUE),sd(as.numeric(ACC_pc),
 na.rm=TRUE),sd(as.numeric(ACC_nodelink),na.rm=TRUE),'\n',sep="\t\
 t");
cat('PREC-
 mean',mean(as.numeric(PREC_tabular),na.rm=TRUE),mean(as.numeric(P
 REC_pc),na.rm=TRUE),mean(as.numeric(PREC_nodelink),na.rm=TRUE),'\
 n',sep="\t\t");
cat('PREC-
 sd',sd(as.numeric(PREC_tabular),na.rm=TRUE),sd(as.numeric(PREC_pc
 ),na.rm=TRUE),sd(as.numeric(PREC_nodelink),na.rm=TRUE),'\n',sep="
 \t\t");
cat('R-
 mean',mean(as.numeric(R_tabular),na.rm=TRUE),mean(as.numeric(R_pc
 ),na.rm=TRUE),mean(as.numeric(R_nodelink),na.rm=TRUE),'\n',sep="\
 t\t");
cat('R-
 sd',sd(as.numeric(R_tabular),na.rm=TRUE),sd(as.numeric(R_pc),na.r
 m=TRUE),sd(as.numeric(R_nodelink),na.rm=TRUE),'\n',sep="\t\t");
cat('F-
 mean',mean(as.numeric(F_tabular),na.rm=TRUE),mean(as.numeric(F_pc
 ),na.rm=TRUE),mean(as.numeric(F_nodelink),na.rm=TRUE),'\n',sep="\
 t\t");
cat('F-
 sd',sd(as.numeric(F_tabular),na.rm=TRUE),sd(as.numeric(F_pc),na.r
 m=TRUE),sd(as.numeric(F_nodelink),na.rm=TRUE),'\n',sep="\t\t");
}

getTime <- function(id,resultType) {
#id is the participant ID, resultType takes either 't', 'p', or
 'n' for tabular, parallel coordinates, or node-link, respectively
#example usage: getTime('101M','t')$'duration'

if (resultType=='t'){fileAppend='-table';}
else if (resultType=='p'){fileAppend='-pc';}
else if (resultType=='n'){fileAppend='-nodelink';}

concat_file_ID =
 paste(sub_directory,id,"/id",id,fileAppend,".txt",sep="");
fileScan=scan(concat_file_ID,what="raw",quiet="TRUE",sep=" ");

if (nchar(fileScan[1])>2){
timeStart=fileScan[pmatch("start:",fileScan)+1];
timeEnd=fileScan[pmatch("end:",fileScan)+1];
timeDuration=as.numeric(difftime(as.POSIXct(timeEnd,format='%H:%M
 :%S'),as.POSIXct(timeStart,format='%H:%M:%S'),units='mins'));
timeDuration = format(round(timeDuration,3),nsmall=3);


result<-
 list("start"=timeStart,"end"=timeEnd,"duration"=timeDuration);
} else {
result<-list("start"="NA","end"="NA","duration"="NA");
}

return(result);


}
```

INTENTIONALLY LEFT BLANK.

Appendix D. List of Post-Task Free Response Comments

This appendix appears in its original form, without editorial change.

Post-task free response comments are presented in this appendix and are unedited. Duplicate responses from a given participant and responses of less than three words are omitted and do not appear in the results.

C.1 Analysts

“What components of the visual displays (table, parallel coordinates, and node link display) were most effective?”

- “I believe tables because the parallel coordinates were a little confusing to me. Maybe after I became used to the view it would be different.”
- “The emphasis of traffic patterns.”
- “Ability to quickly view multiple connections to other nodes/IPs.”
- “All of the components are useful and effective when used properly.”
- “The table was the most effective means of organizing and visualizing an overview of all the data. It was like a spreadsheet.”
- “The table’s ease of use”
- “Correlation of possible malicious activity to port and destination.”
- “Table is much better at identifying the actual alerts. Parallel is more involved with showing all the alerts that are on the node. Node Link is best at showing all the alerts that are attached to each ID”
- “Birds eye view of the situation”
- “Sorting and filtering aspect of the table was most effective in quickly understanding what was happening on the network.”
- “On the table the ability to use parameters to filter the alerts was useful. The thickness of the lines showing traffic levels on the other two were helpful.”
- “The table displays which allowed for viewing the source, destination, port and alert all on one line was helpful.”
- “Node link the visual helped with the flow of traffic.”
- “Table most effective followed by node link”
- “Selecting specific traffic in question.”
- “The table was the most effective. It is easy, you don’t need all these fancy visual tools to find issues. Make everything simpler. The Node display and

the Parallel display looked like a lot of noise. I wasn't interested in using them."

- "Sorting by source or destination location and port. Sorting Time. Ease of viewing connections from left to right clearly."
- "I think the parallel coordinates display was the most helpful to me. I could quickly identify the most active areas. I think this would be a great tool to use in conjunction with the table display."
- "Pop up windows were useful for alerts"

"What aspects of the visualizations (table, parallel coordinates, and node link display) did you like best?"

- "I thought the node link display was interesting. It was nice to see a view of a network topology and the path the data travels."
- "The emphasis of traffic patterns to identify large vs small amounts."
- "Visually separating the internal hosts from external hosts to quickly see flow of data between internal only and internal to external."
- "When used to represent trend analyses and case studies, they are strongest."
- "The table was easy to manipulate and visualize, I could see multiple items at once."
- "Table is much better for seeing and lumping the alerts together. Parallel showed the flow a lot better. Node link showed the flow and other ID numbers to the alerts faster."
- "Of the node link display, I liked how you could visually see external nodes versus internal nodes and how many connections were going to each node."
- "Associations with the various nodes in the node-link."
- "The specific alerts being displayed"
- "Node display was okay but you need to indicate more than just the thickness of a line to determine if there is an incident. Color coding alerts on nodes might be helpful. Breaking the nodes down to a close up view might be more useful."

- “I liked nothing about them, they were convoluted, didn't display information in any type of intelligent fashion. Someone needing to know any information would get to the data needed slowly.”
- “I liked being able to quickly identify the most active times of the day and the mst common request domains of the parallel coordinates displays.”

“What aspects of the visualizations (table, parallel coordinates, and node link display) did you not like?”

- “Parallel coordinates just because it looked foreign to me at the moment.”
- “It takes a while to glance at everything when you can glance at percentages and numbers. Those are quicker to grasp sometimes.”
- “Difficult to view underlying data elements (alerts, ports)”
- “The line display and node display were convoluted at best; they detracted from the information presented. In terms of investigative procedure, I believe they would only serve to stifle.”
- “The parallel coordinates was a nightmare to visualize and the node link display was far too time consuming to hover through.”
- “The way the connections are shown.”
- “Table it is hard to see the relation of the alerts harder, and if there are more than one alert type on the node it can be harder to see that. Parallel, has too much going on that it becomes really lost and confusing, this is definitely not a set up I would want on a larger network. I spend most of my time going through the network trying to make sure that I haven't skipped a node. Node link, has a good basis, but to search quickly it can be difficult if there are more ports added in. This version could use the nodes in it not just the IDs.”
- “Connections were very hard to follow, information was displayed in a non-intuitive manner, correlations were very difficult to find without excessive work.”
- “The parallel coordinates interface was not useful or intuitive and I could not sort the coordinates. Moreover, once I was finished with an alert it should have been removed from my view. Also I should be able to remove noise from my view with a filter. The node link display was slightly more useful but the node sizes were not intuitive nor were the most important piece of analysis data displayed up front: the alerts!”

- "The parallel coordinates was too noisy and visually confusing. In the node link alerts were not readily viewable."
- "I really didn't like the parallel coordinates visualization. I didn't feel like it told me anything about whether a system was compromised or not."
- "Parallel coordinates is good for fine analysis but not for raw / bulk analysis"
- "All the lines of connections, it was somewhat confusing."
- "The graphical representations were completely unusable to me. The table was fine but there needs to be drill down options to see more data. I look at an alert then I check some traffic if I see something suspicious I dump the traffic and do a thorough investigation."
- "Everything, disorganized and overly complicated."
- "I was not a big fan of the node display. It just didnt seem as intuitive. I liked having the data displayed in front of me like on the table and parallel displays."
- "Parallel linking was a little confusing"

"What aspects of the visualizations (table, parallel coordinates, and node link display) helped you to identify intrusions?"

- "Tables because that is the format I am used to."
- "Sometimes hosts generating a lot of traffic and here and there random beacons (very little traffic)"
- "Amount of links and nodes interacting with hosts"
- "The table was intuitive and permitted me freedom to find relations within the given data set. The others were difficult to read and identify events."
- "The type of alert generated, otherwise nothing can be determined without packet capture data analysis as most would be false positives on an enterprise network."
- "Mostly table, but also parallel somewhat."
- "I used prior knowledge of network scanning"
- "Being able to filter out ports and hosts helped on the table. On the parallel coordinates, having a list of the alerts on the right side helped."

- “The table led me to identify the most intrusions because everything is on one line and is easy to look at it. It is also what I am used to.”
- “Ability to group data types”
- “Table data is good but it was lacking. Mainly because the data that was being used. I don't look for MiddleEarth, I look for CH, RU, UK. Country data along with current exploits or know malware, src/dst ports, etc.”
- “Mostly alerts, ports and timing of the attacks and the method involved. The graphical displays however were overly complicated and did not help identify much, other than seeing the direction the data traveled to and from.”
- “Having the most common domains requested displayed like in the parallel coordinates display I think was really helpful. If the most common domain that is being request is malicious it helps paint the big picture of what is going on.”
- “Tabular columnar listings and node pop ups were helpful”

“What features of the visual displays were most useful for completing your tasks in this study?”

- “Potential exploit, dst port, and country.”
- “Legibility, and keeping things simple”
- “Tabular, textual and symbolic shapes.”
- “I liked the features of the table best. It enabled me to read off data much more quickly.”
- “The tabular data was the most helpful.”
- “Visual flow of traffic”
- “Ability to see the data in question”
- “Selecting the alerts and commenting on them.”
- “Table data was fine but I need more than play data to do proper analysis. It isn't just the alert or a darker line of traffic that determines infection or compromise.”
- “Sorting in general whether it be by alert, port, source or destination etc.”
- “The ability to display the big picture.”

“What did you learn from this study?”

- “I didn’t feel it was useful”
- “Its good that I got a chance to see what type of tools can be deployed in future and felt very good to leave feedback about these tools.”
- “Other ways to represent and view information.”
- “That there are some great relationship tools for network intrusion, and some not so great ones.”
- “Being able to see correlations is very important (time, src ip, dst ip, etc). All of this information is very difficult to fit into a graphical interface. Without being able to sort and filter, this makes the analysts job far more difficult. The graphical displays seem decent for a "birds eye view", but a nightmare for actual everyday analysis.”
- “Aggregated data obfuscates signal!”
- “It helped me understand alternative ways of analyzing and detecting network intrusions.”
- “New ideas for looking at data”
- “A simple table can be the better option sometimes.”
- “A tabular-nodal hybrid would be beneficial.”
- “That there are multiple ways to display the same data.”
- “Graphics helps make analysis easier”
- “That there are many different ways of looking at traffic.”
- “With all due respect regarding this project, it seems like there is a long way to go before a very useful graph or node visualization would be effective or efficient to use.”
- “I do not like graphical displays while doing analysis. I find them highly unnecessary, unless of course i was presented with ones that are more intuitive.”
- “Different visualization techniques”

C.2 Students

"What components of the visual displays (table, parallel coordinates, and node link display) were most effective?"

- "Table- the variety of filter options. Node-link- the simple graphics of it"
- "The unique way the data was portrayed gave the data an interesting yet effective way to show data"
- "It was clear and understandable"
- "Ease of following the display itself"
- "The node link displays"
- "Filtering the tables, the thickness of lines in the nodal display"
- "Sorting and alerts information"

"What aspects of the visualizations (table, parallel coordinates, and node link display) did you like best?"

- "The way the different components of the parallel coordinates connected and analyzed"
- "The thickness of the nodal display"
- "Tabular, easy to sort"

"What aspects of the visualizations (table, parallel coordinates, and node link display) did you not like?"

- "The parallel coordinates seemed to be a bit congested with the words"
- "The unsureness of what certain things meant i.e. thickness of links, etc."
- "The table was a bit bland in comparison to the other two displays"
- "Nothing that I did not like"
- "The multiple lines of the parallel coordinates"

"What aspects of the visualizations (table, parallel coordinates, and node link display) helped you to identify intrusions?"

- "Being able to see the amount of traffic and basic information in the node link display helped. Obviously the search options for the table were straight forward and helpful"
- "Noticing clusters of alerts or intrusions"

- “Looking at uncommon activity if it was particular names and times, or times and amount of activity from other networks, etc.”
- “Suspicious activity, mainly instinct and prior knowledge”
- “the thickness of the lines in the nodal link display, the filtering feature of the table”

“What features of the visual displays were most useful for completing your tasks in this study?”

- “The ability to sort and sift”
- “The good organization of the data”
- “I feel as though the line display was the most useful display. It allowed for the user to find alerts easier.”
- “The simple design and colors”
- “Line & bar graphs”
- “Alerts Description and Connections”

“What did you learn from this study?”

- “Advancements in computer vision can greatly impact one's ability to perform a task. Not knowing much about what to look for in this exercise was not a hinderance.”
- “Risks can be categorized based on features that may exist. Network threats are real and something that is really trying to be prevented. Humans must being of some great importance over machines or of equivalent importance when it comes to analyzing threats, otherwise this study wouldn't be conducted”
- “I learned a lot about how the cyber security field operates to try and locate potential computer hazards/viruses.”
- “How to analyze data and pick out abnormalities through a huge amount of data”
- “I learned how to analyze data in an ordinarily factor.”
- “How to detect potential cyber bullies via networks”
- “I learned different ways to visually analyze and look at data”
- “How to find suspicious data”

- “How to analyze data through different data representations”
- “Couldn't view any of the picture so .... may be someone need to improve the link”
- “Different Methods of detecting intruders”
- “I learned how to detect intrusions and attacks”

List of Symbols, Abbreviations, and Acronyms

ARL	US Army Research Laboratory
CND	computer network defense
CNDSP	computer network defense service provider
CSV	comma-separated value
CTA	cognitive task analysis
FN	false negative
FP	false positive
HTML	hypertext markup language
IDS	intrusion detection system
IP	Internet Protocol
MSU	Morgan State University
PC	parallel coordinate
TN	true negative
TP	true positive

1 (PDF)	DEFENSE TECHNICAL INFORMATION CTR DTIC OCA
2 (PDF)	DIRECTOR US ARMY RESEARCH LAB RDRL CIO L IMAL HRA MAIL & RECORDS MGMT
1 (PDF)	GOVT PRINTG OFC A MALHOTRA
2 (PDF)	DIR USARL RDRL HRA AA C J GARNEAU RDRL ROI C R F ERBACHER